University of Nevada, Reno

Decision Making Under Risk and Uncertainty

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

by

Florina Salaghe

Dr. Thomas Harris/Dissertation Advisor

May, 2017

UNIVERSITY

OF NEVADA THE GRADUATE SCHOOL

RENO

We recommend that the dissertation
prepared under our supervision by

Florina Salaghe

entitled

Decision Making Under Risk and Uncertainty

be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

Thomas Harris, Ph.D., Advisor

Mark Nichols, Ph.D., Committee Member

Federico Guerrero, Ph.D., Committee Member

James A. Sundali, Ph.D., Committee Member

Gregory Stone, Ph.D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

May, 2017

Abstracts

Chapter 1 -An Empirical Investigation of Wagering Behavior in a Large Sample of Slot Machine Gamblers

Positive and negative autocorrelated behavior arising from hot hand and gambler's fallacy beliefs has been investigated in many domains such as sports, financial markets, foraging, gambling etc. Using non-experimental, individual transaction gambling data, we examine the existence of positively or negatively autocorrelated betting behavior in a panel of 42,669 gamblers observed over a period of 108 consecutive days encompassing over 17 million slot machines plays. The statistical analyses suggest that gamblers increase the amount bet after wins and decrease the amount bet after losses, on average. These findings remain significant after proxy variables are introduced to control for other competing hypotheses such as house money effects, gambling excitement, a "heavy gamblers" effect, and rare events. Although we are not able to observe players' beliefs directly and cannot know all the factors driving gambler behavior, the findings that slot machine gamblers increase the amount bet after wins and decrease the amount bets after losses is most consistent with the conclusion of a hot hand effect in slot machine gambling.

Chapter 2 - Nevada Ranchers attitudes towards the Trichomoniasis Vaccine - Survey Results

The *Tritrichomonas Foetus* vaccine, developed by University of Nevada in cooperation with Ford Dodge Laboratories, has been available for over twenty years to Nevada cattle producers. The rates of adopting the vaccine are still lagging while the disease incidence in the state is increasing, raising concerns of industry leaders and local authorities. A generalized ordered logit model is employed to find the factors and characteristics that influence the decision making process of Nevada cattle producers regarding vaccination and other alternative public land management practices. Subjective risk attitudes are incorporated in the adoption model and probabilities of adoption for three different groups of respondents are estimated. Results indicate that familiarity with disease treatment, likelihood of exposure and the degree of optimism regarding ranch profitability influence the probability of adoption.

Chapter 3 - An Analysis of Firm's Relocation and Expansion Decision Using Survey Data

Understanding what drives firms' relocation and expansion decisions is critical in formulating sustainable regional economic development policies. Using individual firm level survey data, the current study analyzes factors that influence past and future relocation and expansion decisions by combining the neo-classical, behavioral and institutional approach. Information on over 2,500 firms from different industrial classifications was collected, and both conventional, e.g. economic and physical infrastructure, and unconventional factors, e.g. quality of life and business climate indicators, were incorporated using a self-reporting framework. Results point to internal, external and spatial factors as important predictors of the past and future relocation and expansion decisions and industrial sectors are compared.

Dedication

To my parents, Andronic and Floarea, who taught me that good things are worth fighting for. You are my inspiration every day and I am forever grateful and proud to be your daughter. Thank you for all the love and support and above all, for teaching me the true values in life: hard work, grit and passion to follow my dreams.

Acknowledgments

I still remember the excitement of the first day of graduate school and the unmistakable feeling that I am one step closer to fulfilling my dream: getting a graduate degree in Economics. I had no idea what I wanted to do with my degree but I felt that I was in the right place at the right time. Little did I know that in the mix of enthusiasm, determination, and a touch of fear I would discover my true passion: teaching. It has been an incredible journey and I am grateful to everyone that guided and supported me along the way: my family, friends and professors.

My parents, Andronic and Floarea, for teaching me that it is ok to fail as long as you get up and try again. You taught me that it is ok to dream of impossible dreams and so I dared to reach for them. I feel truly blessed to have you as parents and I thank you for all the love, support, encouragement, and understanding.

My brother, Andronic Florin, for all the love, care and support. You have shown me what unconditional love is and you were, are, and always will be the best big brother in the world.

My friends, old and new, that made me feel that I have a home away from home. Ana, you have always been there for me, every step of the way, and I am forever grateful to have found a sister so far away from home. Thank you for everything you have done for me and for all the wonderful moments we shared throughout the years. Dimitra, Diana and Ilias, I am so grateful to have you in my life and to have shared this journey with you. You have inspired and encouraged me, thank you for having more faith in me than I had in myself. I feel truly blessed to call you all family!

My colleagues and friends at the University Center for Economic Development, Sarah, Marie, and Malieka, I have learned so much from you over these years. You have all been there for me and made this journey so much easier! Thank you for your kind hearts, encouragement and support!

Last but definitely not least, I am forever grateful to the wonderful professors that I had the privilege of working with, I could not have asked for better mentors. My advisors, Tom Harris and Mark Nichols, who helped me navigate these uncharted waters. They always believed in me, encouraged and supported me. My committee members, Federico Guerrero and James Sundali, for all the advice and moral support. Thank you all for helping me see the light at the end of the tunnel and for bearing with me in the darkest moments. I am overwhelmed by your care and determination to help me succeed.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

List of Figures

Chapter 1

Chapter 1

An Empirical Investigation of Wagering Behavior in a Large Sample of Slot Machine Gamblers

Abstract

Positive and negative autocorrelated behavior arising from hot hand and gambler's fallacy beliefs has been investigated in many domains such as sports, financial markets, foraging, gambling etc. Using non-experimental, individual transaction gambling data, we examine the existence of positively or negatively autocorrelated betting behavior in a panel of 42,669 gamblers observed over a period of 108 consecutive days encompassing over 17 million slot machines plays. The statistical analyses suggest that gamblers increase the amount bet after wins and decrease the amount bet after losses, on average. These findings remain significant after proxy variables are introduced to control for other competing hypotheses such as house money effects, gambling excitement, a "heavy gamblers" effect, and rare events. Although we are not able to observe players' beliefs directly and cannot know all the factors driving gambler behavior, the findings that slot machine gamblers increase the amount bet after wins and decrease the amount bets after losses is most consistent with the conclusion of a hot hand effect in slot machine gambling.

1. Introduction

"Present events are connected with preceding ones by a tie based upon the evident principle that a thing cannot occur without a cause which produces it"(Laplace, A Philosphical Essay on Probabilities) .

While individuals may understand the concept of randomness in the abstract, they also sometimes form biased judgments and make erroneous decisions when experiencing randomness. For example, most people will likely agree that a flipped fair coin has a 50-50 chance of being heads. Yet, if five heads are flipped in a row one may judge that tails is "due" or believe that the coin is hot for heads. For a variety of reasons many random events may not appear random to the human brain.

The study of randomness in gambling behavior has a long history in science and literature (Dostoevsky, 2010; Laplace, 1902). Laplace provides the first published account of the "gambler's fallacy" (in 1796), the belief that after a streak of red on the roulette table, black must be due (Ayton & Fischer, 2004). The complement to the belief in negatively autocorrelated random outcomes is the hot hand, a belief in positively correlated outcomes, such as a craps player getting hot and rolling a streak of 7's (Gilovich, Vallone, & Tversky, 1985).

While the term gambler's fallacy clearly implies a bias arising amongst gamblers, forming biased judgments and making erroneous decisions in the face of randomness can exist in many domains. For example, Gilovich, Valone, & Tversky (1985) found there is no "hot hand" in basketball shooting while Camerer (1989) and Brown & Sauer (1993) find evidence that hot hand beliefs do affect prices in basketball betting markets. Diacon & Hasseldine (2007) find that investors choose funds based on past performance, and the gambler's fallacy can partially explain why investors sell winners too soon and hold losers too long (Weber & Camerer, 1998). Chen, Moskowitz, & Shue (2016) found that judges are more likely to grant asylum to an immigrant if they have just deported another one in a previous case, and also found biases in loan officers and baseball umpires. Simmons, Wicklund and Levie (2014) show that entrepreneurs are less likely to pursue a business venture if they have previously failed, and Eggleston and Laub (2002) suggest that young criminals who have early initial successes are more likely to become career criminals.

Since the study of biased beliefs in the face of randomness is obviously a popular topic in academia, a fair question to begin with is why do we need another paper on the hot hand/gambler's fallacy in the gambling domain? Our paper has some unique advantages. First, our detailed data set of slot machine plays has significant methodological advantages over prior studies. Second, the study of slot machine betting patterns, which often occur very quickly and without much forethought, provide interesting insights on evolutionary and learning aspects of hot hand/gambler's fallacy behavior. We expand on these two justifications next.

Many prior hot hand (HH)/gambler's fallacy (GF) studies have been conducted in the laboratory where it is difficult to control subject motivation. A significant advantage of gambling data is that gamblers bet their own money and begin and end their play of their own volition. While there have been previous field studies of gambling behavior, most have lacked sufficient data to adequately control for confounding factors. Since gambling data does not come with a measure of a player's judgment prior to placing a bet, only a conditional hot hand or gambler's fallacy effect can be teased out as a residual, and that requires additional data to control for confounders (e.g., house money effect, income effect) with the potential to induce autocorrelated behavior in players' wagers.

For example, Croson & Sundali (2005) use field data from roulette games to study the existence of HH/GF behavior. Their data did not include detailed measures of the amount bet or income effects. As such, they suggest that one of the possible explanations for the finding of positive autocorrelation in betting behavior could be that players are increasing their bets after wins because they are playing with house money and not because they form erroneous beliefs with respect to the autocorrelation of outcomes. However, given the nature of their data, they did not have the means to address it. Narayanan & Manchanda (2012) also use slot machine data to examine GF/HH behavior. But their data are aggregated by machine, and individual bet-by-bet data on each slot machine could not be identified and thus individual level controls could again not be created.

Our contribution to the existing literature is to examine whether a conditional gambler's fallacy and hot hand effect can be teased out from a large data set of more than 17 million individual player slot machine wagers. We use bet-by-bet data for each player, with every wager being observed alongside a host of other variables allowing us to construct a rather comprehensive vector of potential control variables for autocorrelated behavior in players' wagers. To our knowledge this is the most comprehensive gambling data set ever used to examine gambler's fallacy and hot hand behavior, and gives us a unique opportunity to examine many hypotheses that have been suggested, but have remained untested, in the literature.

To be clear, our field data is not perfect. For example, it has the disadvantage that a few critical variables are unobservable (e.g., players' actual beliefs) and, as with all non-experimental data, manipulation of key parameters is not possible. However, we believe that analysis of this data set provides a unique and significant contribution to the HH/GF literature by allowing us to control for various confounding factors in a highly disaggregated data set over a large number of gamblers, thereby attacking some of the hurdles in such research.

A second justification for this paper is that slot machine play may be related to some evolutionary arguments regarding HH/GF beliefs. Modern slot machine gamblers place their bets and "spin" the wheel very quickly, generally with around two seconds between plays. Such quick play does not lend itself to conscious rational thought and is more likely taking place with automatic cognitive processes. The question then arises as to what default mechanisms are controlling this automated decision making.

Ever since the studies by B. F. Skinner on positive reinforcement (Skinner, 1938, 1989) psychologists have known that operant conditioning is a powerful mechanism that leads animals to repeat a behavior that has been rewarded. Historically, human survival may have depended on an ability to observe patterns, with misjudging a random pattern as non-random having less deadly consequences than the reverse. For example, misperceiving random spots in the bushes as a Cheetah is less deadly than perceiving a Cheetah's spots as random (Haselton, et al., 2009). The human mind, specifically its primal brain, has evolved to solve specific evolutionary problems such as finding mates and reproduction, avoiding getting killed or contracting diseases, and achieving familial protection, among others. Specifically, (Wilke & H., 2009) Wilke & Barret (2009, p. 163) suggest that "the hot hand is an evolved default assumption" that has been learned from foraging in a natural environment of clumped resources. Our data can be used to explore these hypotheses.

The rest of the paper is organized as follows. We begin with a brief survey of the gambler's fallacy and hot hand literature, paying particular attention to the phenomena in gambling contexts. We then present a detailed description of our data set and summary statistics. A regression model is then developed and tested followed by a discussion of the results. We reserve the last section of the paper for concluding remarks and future work.

2. Background Literature

This section is intended to provide some background to the hypotheses testing that follows in the next sections, but it is not intended to be a detailed review of the literature. The following paragraphs are a selective mosaic to illustrate the broad implications of the hot hand and gambler's fallacies.

The study of negative recency effects, or the gambler's fallacy, has a long history in two-choice probability learning studies in psychology (e.g. see Jarvik (1951), Anderson (1960), and see Estes (1964) for a review). The study of positive recency effects, or the hot hand belief, began more recently with the very popular basketball shooting study of Gilovich, Vallone and Tversky (1985).

The most common explanation for the gambler's fallacy belief is the representativeness heuristic (Tversky & Kahneman, 1971). The representativeness heuristic arises from a "belief in the law of small numbers" which suggests that a small sample should be representative of a large sample. If a roulette wheel produces a series of five red outcomes in a row, a gambler might view this sample of five red outcomes as not representative of the large sample which should produce an equal number of red and black outcomes. Thus, after seeing five red outcomes a gambler would (wrongly) believe a black outcome is due in the immediate future in order to bring the small sample back in line with the large sample. The representativeness heuristic is a form of cognitive bias and leads to an expectation of too much alternation in a binary sequence of outcomes and hence results in gambler's fallacy beliefs and expectations.

In their famous basketball study Gilovich et al.(1985) argue the representativeness heuristic can also be used to explain the hot hand theory. If the dice thrower at a craps table tosses outcomes of seven for five throws in a row, other players at the table might believe the dice thrower is "hot" and expect another seven to be rolled. A hot dice thrower produces a streak of outcomes that is unrepresentative of randomness leading to rejection that a random process is producing the outcomes and thus the expectation that more sevens will be tossed.

One interesting line of reasoning on the hot hand bias comes from an evolutionary perspective on cognitive biases (Haselton, et al., 2009). These theorists have suggested that the human mind has evolved to solve specific evolutionary problems such as finding mates and reproduction, avoiding getting killed, avoiding diseases, and familial protection. The success of the human race suggests the human mind has adapted quite well to solve these tasks. Sometimes the human mind uses heuristics that are simple, fast and frugal (Gigerenzer, 2000) and that work very well to solve specific evolutionary problems. For example, since food and other necessary resources for survival are often found in clumps, the formation of hot hand beliefs with regard to finding clumped resources can lead to evolutionary success. Building off the ecologically rationality arguments of Gigerenzer (2000), Wilke & Barret (2009) "propose that the hot hand is an evolved default assumption" (pg. 163) that has been learned from foraging in a natural environment of clumped resources. If belief in the hot hand is an evolved default assumption then it is perfectly reasonable to expect slot machine gamblers to place bets in a manner consistent with wins and losses clumped together.

Behavior consistent with either hot hand or gamblers fallacy may arise in other non-gambling and non-sports settings. Studies have found that entrepreneurs are less likely to pursue a business venture if they have failed previously (Ucbasaran, Westhead, Wright, & Flores, 2010). Thus, belief formation also plays an important role in entrepreneurship. For example, there is empirical evidence suggesting that both social and personal beliefs regarding failed entrepreneurs are important factors that determine whether entrepreneurs engage in subsequent startups, following a failure. Simmons, Wiklund and Levie (2014) study the effect of social stigma and information transparency on subsequent behavior of failed entrepreneurs and find that there is a higher likelihood that entrepreneurs will engage in future startups in countries with low stigma and little publicly available information concerning past failures.

In literature analyzing crime, it has been found that an individual's propensity toward crime is enhanced by early success in criminal activity (Eggleston & Laub, 2002). One consequence of early success in criminal activity is an increase in the severity of subsequent crimes and longer criminal careers (Piquero, Brame, & Lynam, 2004). Brezina and Topalli (2012) study the effect of criminal self-reported efficacy and their effect on future decisions and criminal behavior. Offenders' efficacy beliefs, even if they were "seriously biased and distorted", were found to be positively associated with increased risk-taking behavior (Brezina & Topalli, 2012). Brezina et al. (2012) found evidence in support of the hypothesis that perceived criminal efficacy is negatively related to arrest-to-crime ratios and positively related to the likelihood of future offending. Furthermore, their results suggest a negative correlation between self-reported efficacy and criminals' future intent to "go straight." Thus, there is some empirical evidence suggesting that early criminal career success increases the likelihood of consequent criminal behavior and the seriousness of offenses. Piquero et al. (2004) investigate the relationship between the age of first police contact and criminal career length. Their results indicate that "those parolees exhibiting early onset tend to have longer criminal careers that those parolees exhibiting later onset".

While initially considered separately, more recent research suggests that gambler's fallacy and hot hand beliefs may arise from similar processes (Ayton &

Fischer, 2004). Ayton & Fischer (2004), argue that people expect random devices, such as roulette wheels, to produce random outcomes while people, such as a dice thrower, can get hot and produce streaks. Thus a roulette wheel should produce a locally representative sample of random outcomes which lead to gambler's fallacy expectations when a streak occurs, while a dice thrower might get hot suggesting that random processes are not in control of the outcomes and thus produces hot hand expectations of the continuation of the streak. This line of reasoning supports the hypothesis that slot machine players should exhibit gambler's fallacy, since slot machines are random generating devices. Slot machine gamblers will expect a short term streak to reverse and thus, increase risk taking following a streak of losses and decrease risk taking following a streak of wins. The clumped resource argument reviewed above, however, supports the opposite hypothesis, namely that slot machine gamblers will exhibit hot hand beliefs (Blanchard, Wilke, & Hayden, 2014; Wilke & Barrett, 2009).

How should a "rational" gambler play slot machines? Obviously a perfectly rational decision maker should not play slot machines or gamble at all if the objective is to make money. But if a gambler has decided to play slot machines, say for entertainment purposes, and wants to play as rational as possible then what should the gambler do? Since slot machines payoffs are regulated to be completely random then it really doesn't matter what strategy you use. If the payoffs are completely random then any strategy employed will produce the same random outcomes. Since strategy is removed from optimal slot machine play, many slot machine gamblers play the game quite quickly and decisions become almost automatic for the gambler. Thus, if betting decisions are automatic and if hot hand beliefs are a default evolutionary mechanism then it would be reasonable to expect that slot machine betting patterns would be consistent with hot hand beliefs. That is, if the mind is in an automatic processing mode then the decisions that are output will likely be consistent with the evolutionary wiring of the mind. The finding that slot machine gamblers behave in a manner consistent with the hot hand would thus have significant relevance for other automatic cognitive processing tasks such as visual perception or assembly line work.

A hot hand belief might also form if the gambler believes the slot machine has human like qualities. Roney & Trick (2009) provide evidence in a coin flipping experiment that if a subject's attention is focused on the coin they will expect the streak to reverse (gambler's fallacy), but if a subject's attention is focused on the person flipping the coin they will expect the streak to continue (hot hand). It is possible that some slot machine gamblers project human like qualities onto slot machines (anthropomorphism) and if so, would be more likely to believe that a slot machine could get hot (Kim & McGill, 2011; Riva, Sacchi, & Brambilla, 2015). Most modern slot machines are designed with elaborate themes expressed by sound and video that will contribute to a belief that the machine is more than simply a randomization device (Dixon et al., 2014) .

There have been numerous prior contributions investigating the existence of gambler's fallacy and hot hand behavior in gambling field studies displaying mixed findings. For example, Clotfelter & Cook (1991, 1993) and Terrell (1994) report that individuals are less likely to bet on a lottery number that has recently won, which is in line with gambler's fallacy beliefs. Camerer (1989) found that bettors in basketball markets bet with beliefs that teams get hot or cold. Further muddling the story of a consistent effect, Croson & Sundali (2005) and Sundali & Croson (2006), find that some roulette players bet with a streak (hot hand) while others bet against a streak (gambler's fallacy). In the domain of slot machine gambling, Narayanan & Manchanda (2012) found behavior consistent with gambler's fallacy beliefs, but only in the domain of gains.

3. Description of data

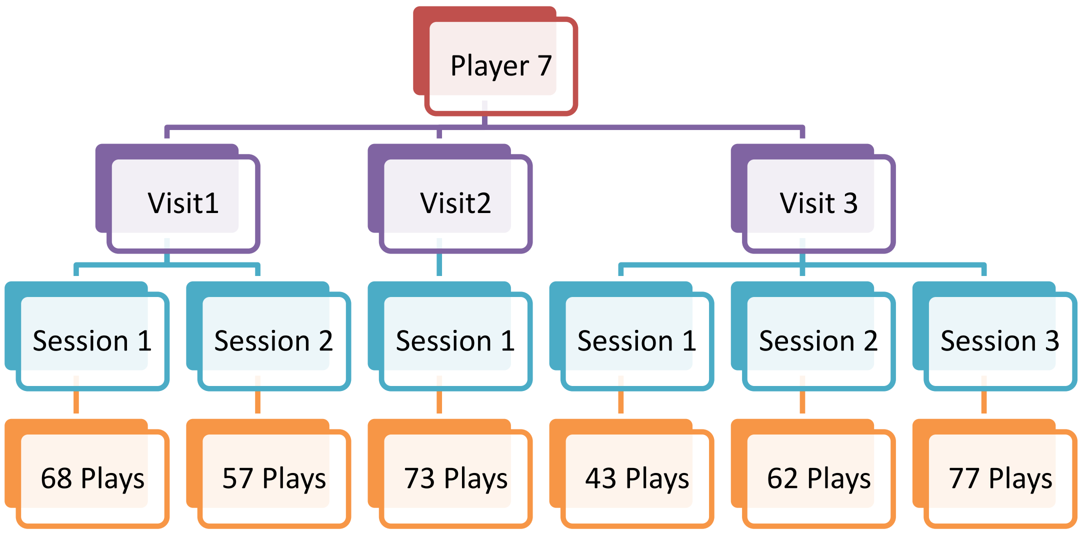
In this paper we use a unique slot machine data set provided by a large designer and manufacturer of slot machines. The company provided a detailed transactional slot machine data set collected over a period of three and a half months, from June 30 to October 15, 2015. Over this period, there is data from 108 distinct days. Two categories of players are observed in the data set, carded and non-carded players. Carded players have signed up with the casino for frequent player benefits such as free rooms, food, or play. Each time a carded player gambles a player card is inserted in the slot machine for player tracking. We confined our analysis in this paper to carded players since it is much easier to identify individuals and also comes with certain demographic identifiers.

A carded players' betting behavior is observed at the highest level of detail, an individual wager. A wager is defined as being a single slot machine pull or press of the "play" button. The data set is a chronological record of a player's betting sequence on a single or multiple machines during the observed period. Some of the variables of interest for the purpose of this study are bet-by-bet information on the amount wagered, the amount lost or won, the machine that was played, the start and end time of a wager, and the age, gender and zip code of each carded player. Three levels of data aggregation are defined:

- **Player level**: data for each player on all days they visit the casino, without differentiating between different days or slot machines. This is the player's sequence of wagers, wins, losses, etc. for the entire June 30th - October 15th period.
- **Visit level**: data for each player are distinguished by the different times the data were collected. Specifically, a new visit is defined as a gap between slot pulls that is longer than 8 hours. Even if chronologically the wagers are observed in the same day, if the player took a break of 8 hours or longer, we consider it a new visit. The majority of the time, a new visit represents a new day the player visited the casino.
- **Session level**: data for each player are distinguished by different sessions within the same visit. A new session is defined by a player moving to a different slot machine or a time gap between bets of at least 30 minutes. The majority of the time, a new session is equivalent to the player moving to a different slot machine.

Figure 1 provides an example of the three levels of aggregation defined above. Player 7 visited the casino three times between June 30th and October 15th and played two sessions in their first visit, one in their second, and three in their third. In the first session of the first visit, they placed 68 bets and in the second they placed 57 bets.

Figure 1. **Hierarchical representation of the data**



The raw data set contains 24,182,076 observations and identifies 46,502 distinct players. Because some of the players wager simultaneously on two different slot machines, they are excluded from the analysis in order to insure that the correct sequence of play can be identified. Non-carded players are also excluded because those players cannot be followed over different sessions and visits. After exclusions, we are left with 42,669 players and 17,042,305 observations (across all players).

3.1 Frequency of Casino Visitation and Wagering

Table 1 below shows some summary statistics for the variables presented in Figure 1. On average, each player makes 1.66 visits to the casino during the time period of the sample data. The total number of distinct visits observed is 70,736 across all players. The average number of sessions is 5.64, while the average number of wagers placed by any player is 399.48.

Table 1. **Summary statistics by hierarchical level**

Variable	Mean per Player	Median per Player	Minimum per Player	Maximum per Player	Total
Visits	1.66	1	1	36	70,736
Sessions	5.64	3	1	302	240,884
Plays	399.48	148	1	34,451	17,045,305

On average, gamblers spend 9.64 minutes on one machine (before they either take a break for longer than 30 minutes or switch machines). The median time spent on one machine is about 6 minutes. The time lapse between two consecutive wagers, which gives an indication of how much time players spend processing the information on the slot machine screen, is 2.71 seconds.

Women place more wagers, 373.67, than men, 364.88. However, while women place more wagers per visit, they have fewer wagers per session, suggesting women place fewer wagers on one machine but change machines more often during a visit than men.

Age also influences wagering. Individuals older than 47, the average age in our sample, visit the casino more often than individuals below the average age (1.73 vs 1.58 visits), play on more machines (3.53 vs 2.57) and place a greater number of wagers (80 vs 69).

3.2 Wagering Expenditures

The average wager amount, without accounting for frequency of play for each player, is $3.94, with men wagering $4.42 on average, while women wager $2.29. As evidenced in Table 2, our distribution is positively skewed. In fact, only 11% of players (4,795) wager above the average amount. The majority of wagers, 53.3% were between

$1 and $5. Wagers less than $1 comprise 37.03% of all wagers. Thus, a total of 90.34% of wagers were under $5 and only 9.66% were above.

Table 2. **Distribution of wagered amounts**

Wagered amount	Frequency	Percent of total	Cumulative percentage
less than $1	6,311,328	37.03%	37.03%
greater than $1 less than $5	9,086,062	53.31%	90.34%
greater than $5 less than $20	1,206,096	7.08%	97.42%
greater than $20 less than $100	390,302	2.29%	99.71%
greater than $100	51,517	0.30%	100.00%
Total	17,045,305	100.00%	

The total amount wagered during the period observed is $67,220,115, with wagers smaller or equal to $1 accounting for 5.57% of the total, while those greater than $100 accounted for 18.86%. Table 3 below details the contributions of various wagering amounts to the total amount wagered. Interestingly, the sum of wagers that are greater than $75 accounts for approximately 23% of the total wagered amount for the period observed in our sample, yet only 1.76% of players have wagers in this category. Thus, 23% of the total dollars wagered comes from less than 2% of the players in our sample. Later in the paper we address this skewness in wagering by separately examining the wagering behavior of gamblers making larger bets.

Table 3. **Wagers amount contribution to total amount wagered**

Wagered amount	total wagered amt	% total wagered amt	cumulative percentage
less than 1	3,745,662	5.57%	5.57%
between 1 and 5	21,743,440	32.35%	37.92%
between 5 and 20	13,362,962	19.88%	57.80%
between 20 and 100	15,689,131	23.34%	81.14%
greater than 100	12,678,920	18.86%	100.00%
Total	67,220,115	100.00%	

Average cumulative losses and wins per player by session, visit, and player are presented in Table 4. Not surprisingly, of course, cumulative losses exceed gains. The average loss at the end of a session is $58.05. Thus, on average, players lose $58 before switching machines. The average loss for each visit is $160.96. For players who win, the average cumulative gain before switching slot machines is $17.92, whereas the cumulative gain per visit is $24.31. The total amount lost by all players during the period analyzed is $41,599,108 and the total amount won by all players is $35,692,320. The difference, casino revenue, is $5,906,788, or 8.79% of the total amount wagered[1].

Table 4. **Player Cumulative Wins and Losses**

	average per session	average per visit	average per player
cumulative win	$17.92	$24.31	$24.25
cumulative loss	$58.05	$160.96	$250.78

4. Empirical Model

As reviewed above, two of the most well documented and studied misperceptions of random sequences are the gambler's fallacy and hot hand. Under the gambler's fallacy, a slot machine player perceives that after a streak of losses you are more likely to see a win (and vice versa). Under the hot hand, a player believes a win or streak of wins will continue. A clear behavioral implication of gambler's fallacy or hot-hand misperception is altering one's wager to account for the expected outcome. A gambler's fallacy player bets against the last outcome or streak. Thus, a win or sequence of wins is

[1] The average theoretical payback of the slot machines in our sample is 90.50%, or an average theoretical hold of 9.5%. Theoretical payback rates, the rate machines are set to pay back on average in the long run, range between 84.5% and 99.54%.

followed by a reduction in one's wager since the predicted probability of another win is lower, whereas a loss or sequence of losses is followed by an increased wager. A hot-hand player exhibits the opposite pattern, betting with the last outcome or streak in the belief that the streak will continue. Thus wagered amounts increase following a win or streak of wins and decrease following a loss or streak of losses. In contrast to a gambler's fallacy or hot hand player, a player that correctly perceives the outcomes as random will not change their wager due to a belief that the next outcome is predictable.

In the empirical model below, we investigate the formation of such beliefs by including indicator variables of whether a player won or lost in the previous plays. One advantage of this data set is that individual wagers and outcomes (win, loss, and amounts) are recorded on each slot machine producing the pattern of wagering at a high level of detail. However, a disadvantage of the dataset is that *beliefs* of gamblers are not observed, only wagers and outcomes are recorded. Therefore, since the hot hand and gambler's fallacy hypotheses are both concerned with *beliefs*, they cannot be tested directly. The empirical strategy is to control for confounding factors that could proxy for autocorrelation in wagers but are unrelated to erroneous beliefs about randomness.

The initial model begins at lowest level of aggregation available in the data set, which is play level data. As previously mentioned, a new play is defined every time a player places a new bet on a given slot machine. That is, a player's sequence of bets is observed as they place consecutive bets on each slot machine during a given visit to the casino. Only accumulated wins and losses during the same visit are allowed to have an effect on the current bet, since we do not observe what happens between consecutive visits. An additional reason to use a within visit specification is that we are primarily interested in investigating belief formation, and the time separation between two consecutive visits limits the probability that the current bet will be affected by events that are temporally spread (Narayanan & Manchanda, 2012).

The following empirical specification is used to model the gambling behavior of the players in our data set:

$$Bet_{ijkt} = \beta_0 + \beta_1 lagbet_{\ ijkt-1} + \beta_2 win3session_{\ ijkt-1} + \ \beta_3 loss3session_{\ ijkt-1} + \theta X_{ijkt-1} + \alpha_i + \varepsilon_{ijkt} \tag{1}$$

Where i represents the player, j represents the visit, k represents the session, and t represents the chronological order of plays.

In the proposed empirical specification, the current period bet is modeled as a function of the previous bet on the same slot machine, with dummy variables for whether the player won or lost in the previous bet, cumulative wins and losses by visit, and a vector of controls that incorporates the confounding factors that might explain autocorrelated behavior (other than the gambler's fallacy and hot hand heuristic). All variables included are detailed below. We also include a constant term, β_0, and individual player fixed effects, α_i, to capture time invariant, unobserved player characteristics that might influence betting behavior and distort our estimates if left unaccounted for.

The variable $lagbet_{\ ijkt-1}$ represents the bet the player places in the previous period, on the same slot machine. The inclusion of this variable accounts for the serial correlation between bets placed on the same slot machine, during the same visit. Omitting this variable would overestimate the response to streaks of wins and losses and underestimate the standard errors of the other independent variables, biasing the results.

The variable $win3session_{ijkt-1}$ is a dummy variable equal to 1 if player i won the previous three bets, during visit j, session k. Similarly, $loss3session_{ijkt-1}$ is a dummy variable equal to 1 if player i lost the previous three bets, during visit j, session k. In other words, $win3session_{ijkt-1}$ is equal to 1 if the player won three times in a row (on the current slot machine) and zero otherwise, while $loss3session_{ijkt-1}$ is equal to 1 if player i lost three times in a row, and zero otherwise. These two variables are included in order to test whether players respond to streaks of wins and losses. A positive coefficient for wins and a negative coefficient for losses would suggest the existence of a positive autocorrelation in bets, whereas a negative coefficient on wins and a positive coefficient losses would suggest the existence of negative autocorrelation in bets. We chose streaks of three wins and three losses following existing literature that suggests that people believe a streak is occurring after the third repeat event in a sequence (Carlson & Shu, 2007).[2] Appendix A contains regression results from the main specification using different streaks. These results are qualitatively consistent and reinforce the idea that players believe in streak formation, as the size of the coefficients increases when we allow for longer streaks.

The average number of three-in-a-row winning streaks per session (per machine) is 0.59. Thus, on average, a player can expect to see 0.59 such winning sequences per session. The maximum number of winning streaks per session, among all players, is 120. Most players, 43.86%, have an average of zero winning streaks per session, and 89.5%

---

2 As a robustness check we estimate the model using streaks of 1, 2, 4 and 5 consecutive wins and losses. The results are qualitatively similar and available upon request.

have one or fewer. Similarly, the same sequence can be observed 1.09 times on average during each visit to the casino. The maximum number of three wins in a row during a player visit is 211. 59.45% of players have an average of zero win streaks per visit, while 80.64% have an average of one or fewer win streaks. There are 126,088 (0.74% of all observations) such winning streaks in the data set. Thus, winning streaks are a rare occurrence in our data set during the time period taken into consideration.

The occurrence of three losses in a row is a lot more common than that of wins. The average number of three losses in a row per session is 35.4. The maximum number of such streaks per session, among all players, is 1,900. By visit, the average number of three losses in a row is 101.53. The maximum number of three losses in a row by visit is 6,795 and the total number of losing streaks, for the entire sample, is 8,375,010 (49.13% of observations).

The vector X_{ijkt-1} contains control variables that differ by specification. For the models that do not include fixed effects, it includes gender and age information for each player. It also incorporates variables that account for factors, other than hot hand and gamblers fallacy, that could be alternative explanations of the observed betting behavior: excitement/utility from gambling (Conlisk, 1993), house money effects (Thaler & Johnson, 1990), and changes in risk attitudes (Kahneman & Tversky, 1979).

In order to account for excitement/utility from gambling we include the gain relative to the amount bet. By including this variable, we attempt to capture the excitement generated by winning relative to the amount bet in each gamble. An alternative way of measuring excitement is to include a dummy variable that is equal to 1 when an unusual win occurs. We define an unusual win as the rare event of winning 5 times the wagered amount during a particular bet. A rare event is generally defined as something that occurs less than 20% of the time (Hertwig, Barron, Weber, & Erev, 2004), which in the dataset is winning 5 times the wagered amount. The standard deviation of payouts over the last five bets is also included to capture the dispersion of the prospect, a measure of excitement suggested by Conlisk (1993). Thus, the higher the standard deviation of payouts, the more excited a player gets since they receive some form of utility from the anticipation of the outcome.

According to Thaler & Johnson (1990), gambling behavior is influenced by previous losses and gains and experimental findings suggest an increased willingness to accept gambles after experiencing gains. This finding is labeled as "the house money effect" and it implies that people play differently when they are winning and are in the domain of gains (playing with house money) than when they are losing and are in the domain of losses (playing with their own money). The house money effect suggests that gambling with house money increases a player's willingness to accept risk. In our model specification, we introduce house money effects and own money effects by interacting a dummy variable of whether a player won or lost three consecutive times with the cumulative amount won or lost for the current visit. Thus, we construct four variables to account for the house money and own money effects: *win3cumwinvisit, loss3cumwinvisit, win3cumlossvisit and loss3cumlossvisit.* Specifically, we are interested in comparing a player's gambling behavior when they win three consecutive times and are playing with house money (*win3cumwinvist*), to when they lose three consecutive times and are playing with their own money (*win3cumlossvisit*). Similarly, we are interested in whether a player responds differently to a streak of three consecutive losses when they play with house money (*loss3cumwinvisit*) and when they play with own money (*loss3cumlossvisit*).

In order to allow players to behave differently in the domains of gains and losses, we include variables *cumwinvisit* and *cumlossvisit* that represent the cumulated amount won or lost respectively, during the current visit. We want to test whether our results are in line with prospect theory, specifically if players wager differently in the domain of gains versus losses. We also include the square terms of cumulative wins and losses in order to allow for a nonlinear relationship between these predictors and the dependent variable, the current period bet.

5. Results

Following the empirical framework presented above, the analysis begins with the simplest specification. In the first specification of Table 5, the amount bet is a function of the previous bet (*Lagbet*), streaks of three consecutive wins and losses (*Wins3session, Loss3session*), and the accumulated wins and losses over the visit and their squares (*Cumwinvist(2), Cumlossvisit*(2)). Results from the OLS specification with clustering at the player level suggest that on average, players increase the amount bet by approximately 5 cents (equivalent to a 1.68% increase in the average wager amount) after experiencing a streak of three consecutive wins (*Wins3session=0.048)*, and they decrease the amount bet by about 2 cents (equivalent to a 0.67% decrease in the average wager) following a streak of three consecutive losses (*Loss3session=-0.016)*. This wagering pattern is consistent with players having hot hand beliefs.

Table 5. **OLS Regression Results with Clustering**

OLS Regression Results

Dependent variable: Bet ij

	(1)		(2)		(3)		(4)		(5)	
	coeff.	p	coeff.	p	coeff.	p	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9876	0.000	0.9876	0.000	0.9877	0.000	0.9864	0.000
Wins3session	0.0485	0.003	0.031	0.069	0.0425	0.006	0.0469	0.004	0.0438	0.007
Loss3session	-0.0165	0.000	-0.0130	0.001	-0.0112	0.003	-0.0149	0.000	-0.0093	0.034
Age	-0.0002	0.091	-0.0002	0.088	-0.0002	0.091	-0.0002	0.090	-0.0002	0.099
Male	0.0152	0.000	0.0152	0.000	0.0154	0.000	0.0152	0.000	0.0134	0.001
Cumwinvisit	0.0001	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.001
Cumwinvisit2	-2.06E-09	0.007	-2.06E-09	0.005	-2.07E-09	0.008	-2.05E-09	0.008	-1.95E-09	0.013
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00010	0.001	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.90E-10	0.003	-5.19E-10	0.002	-4.84E-10	0.004	-4.50E-10	0.005
Win3cumwinvisit			0.00018	0.198						
Loss3cumwinvisit			-0.00006	0.002						
Win3cumlossvisit					0.00003	0.097				
Loss3cumlossvisit					-0.00003	0.000				
Relativegain							0.00095	0.000		
Relativegain2							-3.78E-07	0.013		
Paysd									0.0022	0.010
Rarewin									-0.0326	0.006
Constant	0.0298	0.012	0.0281	0.017	0.0267	0.021	0.0280	0.017	0.0235	0.015
Observations	13,962,769		13,962,769		13,962,769		13,962,769		13,682,392	
R- squared	0.9821		0.9821		0.9821		0.9821		0.9828	

The coefficient for the cumulative amount won and its square term are positive and negative respectively indicating that players increase the amount bet when they are in the domain of gains but at a decreasing rate. This is consistent with prospect theory in that wagering activity is concave in the domain of gains. Our findings suggest that players keep playing when they experience small gains but they cash out when they have accumulated a certain amount. The coefficients for cumulative losses and its square are also positive and negative respectively. Thus, counter to prospect theory, wagering activity is not convex in the domain of losses. Players increase the amount bet after losses, but at a decreasing rate rather than an increasing rate. One possible explanation for the observed behavior in the domain of losses is that players try to "catch up" when they experience losses and thus increase the amount bet in consecutive plays. However, when the amount lost accumulates to a certain point, they give up and cash out. These findings suggest that gambling has an entertainment component, as players increase the amount bet after both wins and losses but only up to a certain point.

Next we address whether players increase the amount bet by more when it comes to gains or losses, or whether the rate of decrease is larger for wins or losses. Results from a Wald test suggest that the coefficient of cumulative losses and gains are not statistically different in this specification[3]. However, we reject the hypothesis that the nonlinear terms are equivalent, meaning that the slopes of the value function are statistically different in the two domains.[4]

---

[3] $F(1, 38207) = 0.98$ Prob > F = 0.3233
[4] $F(1, 38207) = 4.20$ Prob > F = 0.0404

The second specification in table 5 allows for part of the observed betting behavior to be attributed to a house money effect. Players might not necessarily increase the amount bet after wins because they believe they will win again, but instead because they are winning and therefore playing with house money instead of their own. In order to capture this effect, we interact a dummy variable for winning three times in a row with the accumulated wins (*Win3cumwinvisit*=*win3session*cumwinvisit*). The coefficient of the interaction term shows the additional effect of winning three consecutive times when betting with house money. The positive sign of the coefficient (*Win3cumwinvisit=0.00018, p=0.198.*), indicates that players increase the amount bet when they win three consecutive times and are in the domain of gains, although the effect is small and not statistically significant.

Similarly, we include the interaction between an indicator variable for losing three consecutive times with the cumulative gains over the visit (*Loss3cumwinvisit*=*loss3session*cumwinvisit*). The interaction term captures the effect of losing three times in a row when playing with house money (the player is in the domain of gains). This coefficient is negative and significant (*Loss3cumwinvisit=-0.000068, p=0.002),* an indication that players decrease the amount bet after losing three consecutive times even when they are playing with house money, although again the additional impact is small. Interestingly, players' response to streaks diminishes when we control for house money effects, especially when it comes to winning three consecutive times. Specifically, they increase the amount bet after winning, by only three cents as opposed to five, as in the previous specification. Nevertheless, the hot hand effect still persists and is statistically significant. The effect of cumulative gains and losses remains largely unchanged.

The third specification includes the same interaction terms as above, except that now players are in the domain of losses (they are playing with their own money instead of house money). We interact the indicator variable for winning three consecutive times with the cumulative losses for the visit (*Win3cumlossvisit=win3session*cumlossvisit*), in order to capture the effect of winning three times in a row when playing with own money. Results indicate that players increase the amount bet after wins when playing with their own money (they are in the domain of cumulative losses for the visit). We also include the interaction between losing three consecutive times and cumulative losses (*Loss3cumlossvisit=loss3session*cumlossvisit*) in order to capture the effect of losing when playing with own money. Results show that players decrease the amount bet when they are in the domain of losses and experience three consecutive losses. Both coefficients of these interaction terms are statistically significant but quantitatively they do not have much of an impact on players' response to streaks.[5]

The next analysis examines whether winning three consecutive times affects the betting behavior in a different manner when players are playing with their own money (they are in the domain of losses) versus when they are playing with house money (they are in the domain of wins). For this analysis the coefficients of the corresponding interaction terms obtained from the regressions associated with the second and third

[5] Players increase the amount bet after a streak of wins, by 4 cents, and they decrease the amount bet after a streak of losses by 1 cent.

specification are compared. Results indicate that winning[6] or losing[7] doesn't affect subsequent betting differently, regardless of whether the player is in the domain of cumulative wins or cumulative losses.

Specifications (4) and (5) account for the possibility that players get excited after wins and change their betting patterns accordingly. The term introduced in specification (4) to account for excitement is the gain relative to the bet amount (*Relativegain*), together with its square term in order to allow for a nonlinear relationship. Estimation results suggest that this proxy for excitement from gambling has a positive and significant effect on the bet amount. The nonlinear term has a negative sign, indicating that players increase the amount bet when they experience a large win relative to the wagered amount, but at a decreasing rate. This result points to the same conclusion as before, that players increase their bets when they are winning but after a large win relative to the wagered amount, they cash out.

The last specification presented in Table 5 introduces two alternative measure of capturing excitement from gambling (see Conlisk, 1993): the standard deviation of payouts (*paysd*) and an indicator variable for unusual wins (*Rarewin*). Typically, the standard deviation is used as a proxy for risk aversion. Regression results using this specification, show a positive and significant coefficient (*paysd=0.0022, p=0.01)*, an indication that excitement outweighs risk aversion in a gambling setting. Other sources in the literature have documented that gamblers are in general more risk seeking, which is another possible explanation for the positive sign of this coefficient.

[6] Chi2 (1) =1.23   Prob > Chi2 =0.2681
[7] Chi2 (1) =2.68   Prob > Chi2 =0.1015

Rare wins are defined, as previously mentioned, as winning five times the amount wagered. Regression results show that this variable is negative and statistically significant (*Rarewin=-0.0326, p=0.006*), reinforcing the line of reasoning presented above, that players have a tendency of significantly reducing their bets after winning a large amount.

Regardless of the specification presented, our results point to the same conclusion. Players in our dataset seem to respond to streaks by increasing the amount bet after wins and decreasing it after losses. Thus, players are increasing and decreasing bets as if the streak will continue, which is consistent with the hot hand fallacy. This result appears robust to checks of house money effects and excitement from gambling.

6. Robustness checks

Because wining three consecutive times is a rare event (variable *win3session* is equal to 1 in only 0.52% of the cases), we define wins in two other ways in order to increase the occurrence of this event and better assure that a response to a rare event is not driving the observed positive autocorrelation: gross wins and half wins. Gross wins are defined as having any positive coin out amount. A gross win occurs every time a player "wins" part of the amount wagered back, even if the amount won is smaller than the amount wagered. In a similar fashion, half wins are defined every time a player "wins" at least half of the amount wagered. These new ways of defining wins stems from the gambling literature that introduces the concept of losses disguised as wins (Dixon, Collins, Harrigan, Graydon, & Fugelsang, 2015). This literature suggests that players may perceive some monetary losses as psychological wins because of the visual and auditory stimuli produced by the slot machines with partial wins.

Table 6 compares results from three specifications: the net wins specification, where wins are defined as the coin out amount exceeding the wagered amount, the gross wins specification, where wins are defined as any positive coin out amount, and the half wins specification, where wins are defined as a coin out amount equal to at least half of the wagered amount.

Table 6. **Net Wins, Gross Wins and Half Wins**

Dependent Variable: Bet ij	Net_wins		Gross_wins		Half_wins	
	(1)		(2)		(3)	
	coeff.	P	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9877	0.000	0.9877	0.000
Wins3session	0.0485	0.003	0.0403	0.000	0.0431	0.000
Loss3session	-0.0165	0.000	-0.0017	0.725	-0.0127	0.002
Age	-0.0002	0.091	-0.0002	0.066	-0.0002	0.072
Male	0.0152	0.000	0.0156	0.000	0.0147	0.001
Cumwinvisit	0.0001	0.000	0.0001	0.000	0.0001	0.000
Cumwinvisit2	-2.06E-09	0.007	-2.06E-09	0.008	-2.05E-09	0.007
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.84E-10	0.004	-4.84E-10	0.004
Constant	0.0298	0.012	0.0165	0.125	0.024	0.039
Observations	13,962,769		13,962,769		13,962,769	

Results are consistent with those presented in Table 5. Regardless the definition of a win, players increase the amount bet after wins and decrease it after losses. When the definition of gross wins is used, the event of losing three consecutive times becomes less frequent and the coefficient of the indicator variable is insignificant. [8] The coefficients

[8] The event of losing three consecutive times appears in 26.69% of the wagers as opposed to 60.43% of the wagers when using the net loss definition.

for cumulative wins and losses and their squares remain qualitatively the same and changes in magnitude are very small.

Next, the sample is split into two groups of players: those who bet less than the sample average and those who bet more than the average. This analysis examines whether the hot hand fallacy is observed is both groups of players and, if so, to test the hypothesis that "heavy" gamblers respond more to streaks than those who wager smaller amounts. This test will also assess whether the results in previous specifications are driven by "heavy" gamblers.

Table 7 below compares estimation results from the entire sample to those obtained for wagers that are greater than and less than the sample average. In order to make comparisons across regressions more clear a log linear model is used where the dependent variable is the natural log of the amount bet.

Table 7. **Players Comparison by Wager Size**

Dependent variable: ln Bet $_{ij}$

	Full Sample		Wager<$3		Wager>=$3	
	(1)		(2)		(3)	
	coeff.	p	coeff.	p	coeff.	p
Ln(Lagbet)	0.9853	0.000	0.9651	0.000	0.9523	0.000
Wins3session	0.0072	0.000	0.0023	0.045	0.0165	0.000
Loss3session	-0.0067	0.000	-0.0056	0.000	-0.0068	0.000
Age	-0.00004	0.000	0.00004	0.029	-0.0003	0.000
Male	0.0039	0.000	0.0028	0.000	0.0048	0.005
Cumwinvisit	0.00001	0.000	0.00003	0.000	0.00001	0.000
Cumwinvisit2	-2.75E-10	0.000	-1.97E-09	0.000	-1.91E-10	0.000
Cumlossvisit	0.000006	0.000	0.000014	0.000	0.00001	0.000
Cumlossvisit2	-5.06E-11	0.000	-5.58E-10	0.000	-6.13E-11	0.000
Constant	0.0059	0.000	-0.0117	0.000	0.1090	0.000
Observations	13,682,392		10,600,979		3,361,790	
R- squared	0.9828		0.9442		0.9579	

Results are consistent with prior findings. There is also support for the hypothesis that players respond more to streaks when betting larger amounts. The coefficient associated with winning three consecutive times is almost ten times larger in the third specification as compared to the second specification in Table 7.

Gamblers wagering more than $3 increase their wagers by 1.65% following a streak of three wins versus 0.23% for those wagering less than $3. The results are less obvious when it comes to losses, as the corresponding coefficients are closer in size. However, a Wald test for the equality of the coefficients between regressions shows that the null hypothesis that the coefficients are the same can be rejected at 10% confidence level.[9]

7. Conclusion

The primary contribution of this paper is an assessment of whether slot machine gamblers increase or decrease the size of their bets after winning or losing on prior wagers. The statistical analyses of a unique field data set are very consistent: on average slot machine gamblers increase their bets after winning and decrease their bets after losing. Care was taken to eliminate other obvious competing hypotheses that might explain these patterns in the data such as excitement or utility from gambling, house money effects, and own money effects. While the primary findings hold after including these controls, there is, as always, the possibility that other variables not included in the analyses could explain the results.

9 Chi2(1) =3.78 Prob> Chi2 = 0.0519

The seemingly straightforward conclusion to draw is to suggest that on average slot machine gamblers behave in a manner consistent with a conditional hot hand fallacy effect. Drawing on the representativeness heuristic, this would mean that a slot machine gambler expects that a streak of three wins/losses in a row is likely to continue and an increase/decrease in his current bet amount is prudent. To put this another way, we might infer that the gambler is thinking: "If the machine is hot I am going to bet more and if the machine is cold I am going to bet less." The problem with this straightforward interpretation is that our slot machine data set cannot tell us what the slot machine gambler is thinking. What the data set and analyses do tell us is that by far the best prediction of how much a slot machine gambler will bet on the next spin of the wheel is the amount they bet on the last spin of wheel. Thus the safest conclusion that we can draw for the reader is the following: A slot machine gambler will tend to repeat the same bet amount each "spin" and there is a consistent and significant tendency to increase the amount bet after three wins and decrease the amount bet after three losses and this betting pattern is not inconsistent with the hot hand bias. Furthermore, the conditional hot hand effect uncovered is not picking up the positive autocorrelation generated by obvious confounders.

The fact that the data analyses produced a consistent and significant effect is not a trivial result if one has ever observed slot machine gamblers at play. As the regression analyses attest, most slot machine gamblers behave in a fairly robotic manner and simply push the "Spin" button very quickly after the outcome of the last gamble is revealed. What the data analyses suggest is that this robotic, automaton process is altered significantly once a streak occurs and a gambler then makes a change in their bet amount.

An interruption and change in a robotic process is fairly suggestive that a cognitive event has occurred that is strong enough to engage a deliberate neural mechanism that then overrides an automatic processing sequence. Put another way, it would be quite reasonable to expect that an automaton gambler would never systematically vary their bet sizes and that there should have been no systematic and significant reaction to a streak of outcomes and there should have been no evidence to reject the null hypothesis of no effect.

The next question is then what explains gamblers' positive autocorrelation reaction to streaks? We speculate the following. As previously discussed in the context of the representativeness heuristic and in line with the arguments of Ayton & Fischer (2004), gamblers may expect gambling machines, such as roulette wheels, to produce random outcomes but people, such as a dice thrower, can get hot and produce streaks. A slot machine gambler who thus projects of human-like qualities onto slot machines (anthropomorphism) is more likely to believe that a slot machine can get hot (Kim & McGill, 2011; Riva et al., 2015). The evolution of the modern slot machines towards a more personalized, themed version, supports the anthropomorphism hypothesis, and justifies players' belief that a slot machine can be "hot" (Dixon et al., 2015).

The results also contribute to the large literature on prospect theory. The analysis of the cumulative wins and losses during a "visit" suggests that gambling also has an entertainment component: players increase the amount bet after both wins and losses, at least up to a certain point. Although the coefficients of the cumulative gains and losses are quantitatively small, the slopes of the cumulative gains and losses functions are statistically different across specifications, suggesting that slot machine gamblers respond differently in the domain of gains versus losses. Specifically, although players increase their bet in both the domain of cumulative wins and losses, they do so at different rates. Thus, players do respond differently to wins and losses but not entirely in line with the predictions of prospect theory. Our results suggest the shape of the value function is concave in the domain of gains, as predicted by prospect theory, but it appears to also be concave in the domain of losses (in line with the results found by (Vendrik & Woltjer, 2007)), not convex as predicted by prospect theory. Further investigation of this issue is necessary, however, before making definitive conclusions. For example, gamblers may exhibit loss aversion only in the domain of large losses. Average losses per visit might not be large enough to capture loss aversion.[10] Future research should focus on the domain of large losses, as it might be the case that players exhibit loss aversion only above a certain threshold. Another possible explanation is that we have different types of players, for example those that consider the entertainment component as a main driver for gambling versus those that play for financial gains.

The coefficients that capture the streak response get smaller for the specification that includes the house money effects. Results suggest that players increase their wagers after a streak of wins by only 3 cents (as opposed to 5 cents), while still decreasing the amount bet after a streak of losses by approximately the same amount as before (i.e.,2 cents). Thus, we can conclude that the house money effect is clearly noticeable on winning streaks. When accounting for the excitement (disappointment) generated by winning (losing) streaks, our findings suggest that excitement outweighs risk aversion in

[10] The average cumulative loss per player is $160.96 when we do not account for frequency of play

a gambling setting. Results indicate that excitement is positively associated with an increase of the wagered amount in subsequent bets, but players tend to cash out after considerable wins. This result is reinforced by the negative coefficient of the rare win indicator variable that is equal to 1 when a rare win occurs.

Concrete policy implications also arise from our results, specifically in connection with behavioral "nudges" and their effects on problem gamblers. Turner, Liu, and Toneatto (2011) have shown that people with gambling problems are more likely to perceive patterns in random sequences and less likely to identify random sequences than people without gambling problems. Similarly, Orford et al. (2003), summarizing several studies, note that problem gamblers have a greater tendency to misunderstand randomness (e.g., believe random patterns should look random, believe in hot and cold numbers, exhibit the gambler's fallacy) and have a greater belief in their ability to affect an outcome (e.g., belief in luck, illusion of control, and experienced a big win early in their gambling history) than non-problem gamblers. As such, finding evidence of hot hand, gamblers fallacy, or other cognitive biases could serve as a trigger to, for example, remind gamblers that outcomes are random and send information about gambling problem treatment. Gallagher et al. (2014) find that a banner stating "Warning: The Payouts are Random and Not Controlled by Players. Near Wins are Always Losses" resulted in individuals playing less time and decreased informational biases as measured by the Informational Biases Scale (Jefferson and Nicki (2003)) amongst problem gamblers. The warning banner appeared on slot machines before, but not during, actual play. Auer, Malischnig, and Griffiths (2014) show that pop-up messaging on online slot machines in Austria about the length of time an individual has been gambling are effective at inducing a small number of gamblers to cease playing. Other studies, reviewed in Gallagher et al. (2014), have also found that messages about the randomness of games reduces gambling. While we cannot identify whether individuals are problem gamblers, information on the extent of cognitive biases in a large sample of slot machine gamblers has potentially important results for policies designed to mitigate gambling problems or simply better inform consumers about product characteristics.

Future work includes further investigating the lack of a reflection point that we observe in the domain of cumulative gains and losses, a more detailed, individual player analysis, and players' stopping rules. The fact that loss aversion is not present in the context of gambling is a puzzling result and a more detailed analysis might shed some light on this matter. There are a series of competing hypothesis that could explain this finding, such as a change in the location of the reference point as the disconnect between expectations and experience settles in, or the existence of loss aversion only in the domain of large losses. Optimal stopping rules in the context of gambling might also play a role in the lack of convexity of the average player's loss function. Because of the entertainment component of gambling, players stopping rules might differ depending on the size of the bet and the time spent playing.

Other avenues for future work include investigating addictive behavior, recency and saliency of streaks and how they impact betting behavior. As internet gambling becomes more and more popular, there are increasing concerns from industry regulators regarding early identification of risk factors that are good predictors of addictive behavior. Important policy implications and preventive measures can be identified by analyzing addictive behavior in a large sample of slot machine gamblers. Given the great level of detail in our data set, and using literature trends, we can identify the following risk factors as predictors of addictive behavior: trajectory, frequency, intensity and variability of bets (Braverman & Shaffer, 2012). Another topic of interest is investigating the recency and saliency effects and their impact on betting behavior. The study of these effects will not only further our understating of gambling behavior, but also shed some light into other, more general, domains such as insurance and the stock market.

Appendix A

OLS Regression Results
Dependandant variable: Bet $_{ij}$
streaks of 1

	(1)		(2)		(3)		(4)		(5)	
	coeff.	p	coeff.	p	coeff.	p	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9877	0.000	0.9877	0.000	0.9877	0.000	0.9865	0.000
Wins1session	0.0287	0.000	0.0250	0.001	0.0262	0.000	0.0276	0.001	0.0193	0.008
Loss1session	-0.0084	0.038	-0.0093	0.016	-0.0052	0.191	-0.0081	0.049	-0.0079	0.052
Age	-0.0002	0.085	-0.0002	0.084	-0.0002	0.085	-0.0002	0.085	-0.0002	0.093
Gender	0.0150	0.001	0.0150	0.001	0.1520	0.000	0.0150	0.001	0.0132	0.001
Cumwinvisit	0.0001	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.001
Cumwinvisit2	-2.05E-09	0.008	-2.04E-09	0.007	-2.06E-09	0.008	-2.05E-09	0.008	-1.95E-09	0.013
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00009	0.002	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.86E-10	0.004	-4.94E-10	0.003	-4.84E-10	0.004	-4.50E-10	0.005
Win1cumwinvisit			0.00005	0.071						
Loss1cumwinvisit			0.00001	0.568						
Win1cumlossvisit					0.00001	0.389				
Loss1cumlossvisit					-1.69E-05	0.056				
relativegain							0.0003	0.176		
relativegain2							-1.60E-07	0.074		
paysd									0.0022	0.010
Rarewin									-0.0254	0.038
Constant	0.0225	0.0480	0.2394	0.033	0.0205	0.065	0.2223	0.051	0.2126	0.042
Observations	13,962,769		13,962,769		13,962,769		13,962,769		13,682,392	
R- squared	0.9821		0.9821		0.9821		0.9821		0.9828	

OLS Regression Results

Dependandant variable: Bet $_{ij}$

streaks of 2

	(1)		(2)		(3)		(4)		(5)	
	coeff.	p	coeff.	p	coeff.	p	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9877	0.000	0.9876	0.000	0.9877	0.000	0.9864	0.000
Wins2session	0.0302	0.000	0.0234	0.006	0.0289	0.000	0.0281	0.000	0.0260	0.001
Loss2session	-0.0135	0.003	-0.0102	0.017	-0.0085	0.054	-0.1193	0.007	-0.0104	0.027
Age	-0.0002	0.082	-0.0002	0.077	-0.0002	0.080	-0.0002	0.082	-0.0019	0.090
Gender	0.0148	0.001	0.0147	0.001	0.0152	0.000	0.0148	0.001	0.0131	0.001
Cumwinvisit	0.0001	0.000	0.0014	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.001
Cumwinvisit2	-2.05E-09	0.008	-2.12E-09	0.004	-2.07E-09	0.007	-2.50E-09	0.008	-1.95E-09	0.013
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00010	0.001	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.88E-10	0.004	-4.99E-10	0.003	-4.84E-10	0.004	-4.50E-10	0.005
Win2cumwinvisit			0.00008	0.066						
Loss2cumwinvisit			-0.00005	0.067						
Win2cumlossvisit					4.46E-06	0.419				
Loss2cumlossvisit					-2.63E-05	0.000				
relativegain							0.0009	0.000		
relativegain2							-3.62E-07	0.013		
paysd									0.0022	0.010
Rarewin									-0.0355	0.005
Constant	0.0279	0.0180	0.0264	0.020	0.0248	0.032	0.0262	0.026	0.0241	0.019
Observations	13,962,769		13,962,769		13,962,769		13,962,769		13,682,392	
R- squared	0.9821		0.9821		0.9821		0.9821		0.9828	

OLS Regression Results

Dependandant variable: Bet $_{ij}$

streaks of 4

	(1)		(2)		(3)		(4)		(5)	
	coeff.	p	coeff.	p	coeff.	p	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9876	0.000	0.9876	0.000	0.9877	0.000	0.9864	0.000
Wins4session	0.1402	0.001	0.0657	0.159	0.1311	0.001	0.1384	0.002	0.1347	0.002
Loss4session	-0.0159	0.000	-0.0131	0.000	-0.1026	0.000	-0.1459	0.000	-0.0079	0.010
Age	-0.0002	0.082	-0.0022	0.077	-0.0002	0.081	-0.0002	0.082	-0.0002	0.092
Gender	0.0147	0.001	0.0147	0.001	0.0152	0.000	0.0147	0.001	0.0131	0.001
Cumwinvisit	0.0001	0.000	0.0013	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.001
Cumwinvisit2	-2.05E-09	0.008	-2.10E-09	0.004	-2.70E-09	0.007	-2.05E-09	0.008	-1.95E-09	0.013
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00010	0.001	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.88E-10	0.004	-5.09E-10	0.003	-4.12E-07	0.004	-4.50E-10	0.005
Win4cumwinvisit			0.00079	0.048						
Loss4cumwinvisit			-0.00004	0.002						
Win4cumlossvisit					0.00004	0.168				
Loss4cumlossvisit					-0.00003	0.000				
relativegain							0.0011	0.000		
relativegain2							-4.12E-07	0.013		
paysd									0.0022	0.010
Rarewin									-0.0294	0.013
Constant	0.0264	0.0240	0.0259	0.026	0.0242	0.035	0.0250	0.032	0.0211	0.032
Observations	13,962,769		13,962,769		13,962,769		13,962,769		13,682,392	
R- squared	0.9821		0.9821		0.9821		0.9821		0.9828	

OLS Regression Results

Dependandant variable: Bet $_{ij}$

streaks of 5

	(1)		(2)		(3)		(4)		(5)	
	coeff.	p	coeff.	p	coeff.	p	coeff.	p	coeff.	p
Lagbet	0.9877	0.000	0.9876	0.000	0.9876	0.000	0.9877	0.000	0.9864	0.000
Wins5session	0.2088	0.009	0.0501	0.589	0.2064	0.010	0.2072	0.009	0.2045	0.010
Loss5session	-0.0180	0.000	-0.0149	0.000	-0.1191	0.000	-0.0168	0.000	-0.0109	0.000
Age	-0.0002	0.081	-0.0002	0.076	-0.0002	0.080	-0.0002	0.081	-0.0002	0.090
Gender	0.0147	0.001	0.0147	0.001	0.0152	0.000	0.0148	0.001	0.0131	0.001
Cumwinvisit	0.00011	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.000	0.0001	0.001
Cumwinvisit2	-2.05E-09	0.007	-2.11E-09	0.004	-2.08E-09	0.007	-2.05E-09	0.008	-1.95E-09	0.013
Cumlossvisit	0.00008	0.002	0.00008	0.002	0.00009	0.001	0.00008	0.002	0.00008	0.002
Cumlossvisit2	-4.84E-10	0.004	-4.88E-10	0.004	-5.15E-10	0.003	-4.84E-10	0.004	-4.50E-10	0.005
Win5cumwinvisit			0.00149	0.192						
Loss5cumwinvisit			-0.00005	0.003						
Win5cumlossvisit					1.30E-05	0.301				
Loss5cumlossvisit					-0.00003	0.000				
relativegain							0.0011	0.000		
relativegain2							-4.21E-07	0.013		
paysd									0.0022	0.010
Rarewin									-0.0295	0.014
Constant	0.0261	0.025	0.0256	0.028	0.0241	0.035	0.0248	0.033	0.0216	0.03
Observations	13,962,769		13,962,769		13,962,769		13,962,769		13,682,392	
R- squared	0.9821		0.9821		0.9821		0.9821		0.9828	

References

Anderson, N. (1960). Effect of first order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology*(59(2)), 73-93. doi:10.1037/h0049023

Auer, M., Malischnig, D., & Grifiths, M. (2014). Is "pop up" messaging in online slot machine gambling effective as a problem gambling strategy? *Journal of Gambling Issues, 29*, 1-10.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory and Cognition*, *32*(8), 1369-1378.

Blanchard, T. C., Wilke, A., & Hayden, B. Y. (2014). Hot-hand bias in rhesus monkeys. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*(3), 280-286.

Braverman, J., & Shaffer, H. J. (2012). How do gamblers start gambling: Identifying behavioral markers for high-risk Internet gambling. *European Journal of Public Health*, *22*(2), 273-278.

Brezina, T., & Topalli, V. (2012). Criminal Self-Efficacy. Exploring the Correlates and Consequences of a " Successful Criminal" Identity. *Criminal Justice and Behavior, 39*(8), 1042-1062.

Brown, W. O., & Sauer, R. (1993). Does the Basketball Market Believe in the Hot Hand?Comment. *The American Economic Review*(83(5)), 1377-1386.

Camerer, C. F. (1989). Does the Basketball Market Believe in the Hot Hand? *The American Economic Review*, *79*(5), 1257-1261.

Carlson, K. A., & Shu, S. B. (2007). The rule of three: How the third event signals the emergece of a streak. *Organizational Behavior and Human Decision Processes*, 113-121.

Chen, D., Moskowitz, T. J., & Shue, K. (2016). Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Oficers, and Baseball Umpires. *The Quarterly Journal of Economics*, 131(3), 1181-1242.

Clotfelter, C. T., & Cook, P. J. (1991). Lotteries in the Real World. *Journal of Risk and Uncertainty*, *4*(3), 227-232.

Clotfelter, C. T., & Cook, P. J. (1993). Notes: The "Gambler's Fallacy" in Lottery Play. *Management Science, 39*(12), 1521.

Conlisk, J. (1993). The Utility of Gambling. *Journal of Risk and Uncertainty* , 255-275.

Croson, R., & Sundali, J. (2005). The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos. *Journal of Risk and Ucertainty, 30*(3), 195-209. doi:10.1007/s11166-005-1153-2

Diacon, S., & Hasseldine, J. (2007, January). Framing effects and risk perception: the effect of prior performance presentation format on investment fund choice. *Journal of Economic Psychology, 28*(1), 31-52.

Dixon, M. J., Collins, K., Harrigan, K., Graydon, C., & Fugelsang, J. A. (2015). Using Sound to Unmask Losses Disguised as Wins in Multiline Slot Machines. *Journal of Gambling Studies, 31*, 183-196.

Dostoevsky, F. (2010). *The Gambler (Penguin Classics).* London: Penguin. Retrieved from London U6 - ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/summon.serialssolutions.com&rft_val_fmt=info:ofi/fmt:kev:mtx:book&rft.genre=book&rft.title=The+Gambler+(Penguin+Classics)&rft.au=Dostoevsky,+Fyodor,+1821-1881&rft.date=2010-01-01&r

Eggleston, E. P., & Laub, J. H. (2002). The Onset of Adult Offending: A Neglected Dimension of the Criminal Career. *Journal of Criminal Justice, 30,* 603-622.

Estes, W. (1964). Probability Learning. In A. W. Melton, *Categories of Human Learning* (pp. 89- 123). New-York: Academic Press.

Gallagher, T., Nicki, R., Otteson, A., & Elliott, H. (2011). Effects of Video Lottery Terminal (VLT) Banner on Gambling: A Field Study. *International Journal of Menthal Health Addiction, 9*, 126-133.

Gigerenzer, G. (2000). Ecological rationality. *International Journal of Psychology, 35*(3-4), 284-284.

Gilovich, T., Valone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295-314.

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive Rtaionality: An Evolutionary Perspective on Cognitive Bias. *Social Cognition, 27*(5), 733-763.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004, August). Decisions from Experience and the Effect of Rare Events in Riky Choice. *Psychological Science, 15*(8), 534-539.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*(41(4)), 291-297. doi:10.1037/h0056878

Jefferson, S., & Nicki, R. (2003). A New Instrument to Measure Cignitive Distortions in Video LOttery Terminals Users: The Informational Biases Scale (IBS). *Journal of Gambling Studies, 19*(4), 387-403.

Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgement of Representativeness. *Cognitive Psychology, 3*, 430-454.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decisions under Risk. *Econometrica (47)*, 263-291.

Kim, S., & McGill, A. L. (2011). Gaming with Mr. Slot or gaming the slot machine? Power, anthropomorphism, and risk perception. *Journal of Consumer Research, 38*(1), 94-107.

Laplace, P. E. (1902). A philosophical essay on probabilities.

Narayanan, S., & Manchanda, P. (2012). An Empirical Analysis of Individual Level Casino Gambling Behavior. *Quantitative Marketing and Economics, 10*(1), 27-62.

Orford, J., Sproston, K., Erens, B., White, C., & Mitchell, L. (2003). *Gambling and Problem Gambling in Britain.* Brunner-Routledge Press: East Sussex.

Piquero, A. D., Brame, R., & Lynam, D. (2004). Studyng Criminal Career Lenght Through Early Adulthood Among Serious Offenders. *Crime and Delinquency, 50*(3), 412-435.

Riva, P., Sacchi, S., & Brambilla, M. (2015). Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied, 21*(4), 313-325.

Roney, C. J., & Trick, L. M. (2009). Sympathetic magic and perceptions of randomness: The hot hand versus the gambler's fallacy. *Thinking & Reasoning, 15*(2), 197-210.

Simmons, S. A., Wiklund, J., & Levie, J. (2014). Stigma and business failure: implications for entrepreneurs' career choices. *Small Business Economics, 42*, 485-505. doi:10.1007/s11187-013-9519-3

Skinner, B. F. (1938). *The Behavior of Organisms. An Experimental Analysis.* D.Appleton-Century Company, INC.

Skinner, B. F. (1989). *The Origins of Cognitive Thought.* american Psychologyst.

Sundali, J., & Croson, R. (2006). Biases in Casino betting: The Hot Hand and the Gambler's Fallacy. *Judgement and Decision Making, 1*(1), 1- 12.

Terrell, D. E. (1994). A Test of the Gambler's Fallacy: Evidence from Pari-mutuel Games. *Journal of Risk and Uncertainty, 8*(3), 309-317.

Thaler, R. H., & Johnson, E. J. (1990, June). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science, 36*(6), 643-660.

Turner, N. E., Liu, E., & Toneatto, T. (2011). What Does a Random Line Look Like: An Experimental Study. *International Journal of Mental Health Addiction, 9*, 60-71.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76(2)*, 105 -110.

Ucbasaran, D., Westhead, P., Wright, M., & Flores, M. (2010). The nature of entrepreneurial experience, business failure and compartaive optimism. *Journal of Business Venturing, 25*, 541-555.

Vendrik, M., & Woltjer, G. (2007). Happiness and loss aversion: Is utility concave or convex in relative income? *Journal of Public Economics, 91*, 1423-1448.

Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization, 33*, 167-184.

Wilke, A., & H., B. C. (2009, May). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior, 30*(3), 161-169.

Chapter 2

Nevada Ranchers attitudes towards the Trichomoniasis Vaccine - Survey Results

Abstract

The *Tritrichomonas Foetus* vaccine, developed by University of Nevada in cooperation with Ford Dodge Laboratories, has been available for over twenty years to Nevada cattle producers. The rates of adopting the vaccine are still lagging while the disease incidence in the state is increasing, raising concerns of industry leaders and local authorities. A generalized ordered logit model is employed to find the factors and characteristics that influence the decision making process of Nevada cattle producers regarding vaccination and other alternative public land management practices. Subjective risk attitudes are incorporated in the adoption model and probabilities of adoption for three different groups of respondents are estimated. Results indicate that familiarity with disease treatment, likelihood of exposure and the degree of optimism regarding ranch profitability influence the probability of adoption.

1. Introduction

Past surveys of Nevada ranchers show that cattle producers have not adopted the necessary management practices designed to limit the incidence of Trichomoniasis (Trich), a "venereal disease of beef herds caused by the protozoan Tritrichomonas Foetus" (Bhattacharyya, Harris, Kvasnicka, & Veserat, 1997). This disease is common in the Western United States and Florida because of free range commingling of herds on large tracts of public land. Trich is contracted during breeding (from bulls to cows and vice versa) and it does not have any obvious signs, making it difficult to identify and isolate the infected animals in order to prevent spread. When a bull is infected, he will be infected for life and there are no vaccines or treatments available to the cattle producers (Thain, Bruce, & Torell, Trich in Nevada, 2007a). There are however tests that can help identify the infected bulls. These tests are not available for cows. A Tritrichomonas foetus vaccine was developed by the University of Nevada in cooperation with Ford Dodge Laboratories and the USDA granted a conditional license to market the vaccine in 1989. This vaccine is only effective for cows and it helps limit the disease and the loss of cattle crop. Nevertheless, the vaccine only lasts for one breeding season and it is effective in 65% of the cases.

In more recent years in Nevada, rising concerns at the industry level have prompted cattle producers to enlist the help of public regulators in controlling Trich (Thain, Bruce, & Torell, Trich in Nevada. What Other States Are Doing, 2007b). The results of a survey carried out in 2006 revealed that approximately 57% of cow-calf operators in Nevada did not annually test their bulls for Trich and 62% did not use the vaccine for cows and heifers (Thain, Bruce, & Torell, Trich in Nevada. Producer's Concerns and the Financial Impact, 2007c). Thus, the rate of adoption of the vaccine and other Trich management practices remained low while the disease incidence increased among cattle producers. The divergence between industry and individual interests might be the cause of low adoption rates since a free rider problem might arise. Similar to the tragedy of the commons, a rancher's biosecurity actions might influence the behavior of other ranchers that use the same public land. If one assumes that the other cattle producers have good disease management practices, he will be tempted not to test and vaccinate since the probability of the herd contracting Trich decreases. Similarly, vaccinating one's cows may limit the incidence of the disease but does not prevent infection from other ranchers' bulls. This calls for appropriate public regulations that can help control Trich incidence in Nevada and limit the financial losses incurred by cattle producers that use public lands.

The importance of understanding the reasons behind lagging vaccine adoption rates, despite an increase in disease incidence, is reinforced by the magnitude of the financial losses suffered by producers and industry as a whole, which were estimated to be between $3,000,000 and $5,300,000 in 2006, based on survey information analyzed by Thain et al. (2007c). Furthermore, the same survey revealed that 83% of cattle producers in Nevada wanted to see some form of Trich regulation at the state level. These findings set the stage for a more in depth investigation of the issue, in order to voice some important policy implications that need to be addressed by local authorities.

In an effort to better understand Nevada cattle producers' attitudes toward the adoption of the Trich vaccine and other preventive management practices, a comprehensive survey was conducted during 2012 and 2013 by the University of Nevada, Reno Center for Economic Development (UCED). Subjective measures of risk attitudes were incorporated in the survey in an attempt to better understand their effect on ranchers' behavior regarding vaccine adoption. The present study investigates the data collected and an adoption model is developed in order to investigate how ranch specific factors, familiarity with the treatment, individual characteristics, and attitudes towards risk, influence the probability of adopting the vaccine.

A previous study in the state of Nevada investigated the factors that influence the adoption of new technology among Nevada ranchers. Specifically, a similar survey was conducted four years after the vaccine was marketed (1993), in an attempt to identify the factors and characteristics that affect the adoption behavior of ranchers. The study concluded that computer use, veterinary check-ups, herd size and cooperative extension programs are the main factors that influence the probability of adoption (Bhattacharyya, Harris, Kvasnicka, & Veserat, 1997).

The survey data used in the current study allows for a comparison of the results since it has similarly constructed questions. In addition, the present survey incorporates information regarding risk attitudes and includes lottery questions to test prospect theory predictions in the domains of wins and losses. Since the previous survey was conducted in 1993, just a few years after the vaccine was available on the market, it revolved around the factors that enhanced the probability of adoption of new technology. The data for the present study is updated and it allows us to observe the change in the rates of adoption over time and the change in the factors that play an important role in the decision making process, given that the vaccine has been available for over two decades. For example, although computer use played an important role in the adoption of the vaccine when it was first introduced in the market, our most recent survey data indicates that it does not play a significant role anymore, since the majority of ranchers use computers for their daily ranch management operations.

Analyzing the data available from the most recent survey of Nevada ranchers will shed light on the motivation for adopting Trich management practices, including the vaccine. We hope to formulate important policy implications that will help local authorities identify the appropriate measures to take in order to alleviate industry leaders’ concerns regarding Trich.

The rest of the paper is divided into 5 sections. We start by summarizing the survey results, then proceed with the model specification based on literature trends. Next we propose an estimation procedure and motivate the use of a generalized order logit model with different parameter constraints. Finally, results from three different specifications are discussed, followed by an assessment of the goodness of fit and concluding remarks.

2. Survey Results

In an effort to better understand the decision making process of ranchers when choosing disease control management practices, a survey of Nevada cattle producers was conducted by the University of Nevada, Reno Center for Economic Development, during 2012 and 2013. A sample of 800 Nevada ranchers was randomly selected out of 1952 mailing addresses and a questionnaire, a cover letter and addressed stamped return envelope was mailed to every address in the random sample. Out of the 800 questionnaires mailed, 217 surveys were completed. The response rate, computed using the appropriate AAPOR (American Association for Public Opinion Research) formula, was approximately 35.6%[11] (Mahmoudi, Landis, Fadali, & Harris, 2013).

The survey was divided into 9 sections that summarize the main areas of interest:

1) Ranching experience
2) Trich and disease management practices

---

[11] The formula used to calculate the response rate is: $RR = \frac{(I+P)}{(I+P)+(R+NC+O)+e(UH+UO)}$

Where:
RR= Response rate
I=Complete survey
P=Partial Survey
R=Refusal and break-off
NC=Non-contact
O=Other
e= Estimated proportion of cases of unknown eligibility that are eligible
UH=Unknown if household/occupied
UO=Unknown, other

3) Herd history with Trich exposure

4) Trich testing

5) Trich vaccination

6) Attitude towards risk

7) Opinions on Trich regulation

8) Ranch size and management

9) Demographics

The respondents were grouped into three categories based on their attitudes towards the Trich vaccine: Users (ranchers who have adopted the vaccine), Potential Users (respondents who have not adopted the vaccine but declare they might do it in the future), and Nonusers (respondents who have not used the vaccine and declare they will not use it in the future).

Following Bhattacharyya et al. (1997), Table 1 below presents characteristics of respondents by category in order to facilitate intergroup comparison. Summary statistics are presented for both the entire sample and the three categories individually. A discussion of the characteristics of the sample follows.

Table 8: Characteristics of Respondents by Category

	Units	Users	Potential Users	Nonusers	Entire Sample
Survey responses utilized	Numbers	68.00	76.00	45.00	189.00
Herd size	100 heads	7.98	4.77	10.69	7.36
Age	Years	57.21	56.75	56.78	56.92
Education	Years	14.08	13.85	13.87	13.90
Experience	Years	34.18	35.00	34.13	34.50
PC User	%	60.66	46.38	42.86	50.91
Land size	1,000 acres	128.89	94.70	113.19	111.40
Ranch income	% of total income	63.87	46.92	60.21	56.09

Of the 217 respondents that completed the questionnaire, 189 or 87.10% answered the questions referring to the vaccine adoption, while 12.9% skipped them. Respondents were divided into 3 categories as mentioned above. Of 189, approximately 36% of respondents are part of Current Users and 40% belong to the Potential Users group whereas 24% have not adopted the vaccine and say they are not going to do so in the future. Respondents that have not adopted the vaccine claim they do not have a Trich problem and lack the necessary information on the availability of the vaccine.

The average age of the individuals that completed the survey was 56.92 years. Respondents that have adopted the vaccine were on average, slightly older and more educated. Their average age is 57.21 years compared to 56.75 and 56.78 years, respectively. The average years of education for the User category are 14.08, just slightly over the sample mean of 13.9. 77.78% of ranchers that participated in the survey had some specialized training or education, such as technical school training, animal science degree, Cooperative extension education or Cattlemen's update. The highest frequency form of special education/training was Cattlemen's Update, with more than half the respondents reporting attending this event. The average years of experience for ranchers in the sample are about the same across the different groups. It ranges from 3 to 80 years, while the highest average is 35 years for the potential users category. Average land size, comprising of deeded, public, leased and other land, was 111.40 thousand acres for the entire sample. The User category had a higher mean for land-holding size, 128.89 thousand acres, approximately 15 thousand acres larger than the Nonuser category and 33 thousand acres larger than the Potential Users group. A similar trend is seen for the percentage of income that comes from ranching activities, as the importance of ranching income is highest for Users. There is a 3% gap between Users and Nonusers and a 7% gap between Users and Potential Users. On average, 56% of the household income comes from ranching operations.

Together with the land size, the number of animals is another variable that provides information on ranch size and production endowment. The herd is comprised of cows, heifers, steers and bulls. The average herd size was

735.86 heads. The Nonusers category owns the highest number of livestock, followedby Users and Potential Users. This last category has significantly less animals, about 50% less compared to the leading group.

Since the vaccine has been available on the market for more than 20 years when the survey was conducted, we expect the majority of ranchers to be familiar with both the symptoms and the treatment options available to them. 69% of the respondents reported being either familiar or very familiar with the treatment of the disease, and 88% of the 69% were also familiar or very familiar with the symptoms of Trich.

Similar results were found with regards to Trich testing. Vaccinating cows and testing bulls are only two of the large array of disease management practices. As it turns out there are other more popular, but maybe not as efficient, ways for ranchers to deal with Trich. Among these, purchasing bulls that are certified to be Trich negative, annual testing of bulls, fencing and conducting annual pregnancy tests of cows are the most commonly used by respondents in our sample.

Survey participants are also required to specify the reasons for choosing to test their bulls or vaccinate their cows against Trich, or alternatively, not doing so. The top two reasons for testing were maintaining herd health and conception rates, and being responsible. The most frequent reasons for not testing their bull battery were that positive results are reported to authorities, there is no Trich in their ranching area, and the fact that testing is not efficient if neighbors don't also test. When it comes to adopting the vaccine, the most frequent answers for not doing so were the fact that the respondents believed that they did not have a Trich problem and that they did not have enough information on the vaccine. The main information sources, when it comes to ranch management, were the other ranchers followed closely by specialty magazines and newspapers. When it comes to Trich management practices, respondents cite as their primary source of information their vet and other ranchers. 72% of respondents had a herd veterinarian. These findings can offer valuable insight for policy makers, who can focus their limited resources on targeting those channels that are most effective in achieving their goals.

The survey also contains information related to subjective risk measures of contracting Trich. The majority of the respondents believe their herds are at low risk of contracting the disease, relative to other cattle ranches in the state. This low risk perception is cause of concern and it helps explain why Trich management practices are lagging and financial losses caused by this venereal disease are still issues brought up by industry leaders and local authorities. Even though respondents underestimate their own risk of their cattle contracting Trich, 45.81% of them suspect their neighbors' cattle of being infected.

Respondents' general risk attitudes are also incorporated into the survey. The ranchers in the Nonusers category are the least risk averse, while individuals in the other two categories have about the same level of risk aversion, on average. Attitudes towards risk are also influenced by the tendency to worry about unplanned events, such as being hurt or killed in a car accident, being diagnosed with cancer, getting bitten by a rattle snake or getting struck by lightning. Being confronted with these four choices, we find that the most common concern within each group is being diagnosed with cancer. The individuals in the Potential Users category worry most about all of the four events relative to the other groups. The Nonusers worry the least.

In summary, the average respondent in the Users category is slightly older and more educated than the other two groups. Users have the highest proportion of ranchers that use computers for their operations, and have internet at their place of residence. Average land holdings were the highest for this group and they have the highest percentage of income coming from ranch operations. This category is most familiar with Trich treatment and has the highest percentage of ranchers with specialized education/training. It also has the highest percentage of ranchers that test their bulls annually. 93.65% of the individuals in this category list the veterinarian as their primary source of information on the disease. This group also has the highest percentage of individuals that suspect their neighbors' cattle of being infected with Trich in the past five years and they are the most risk averse.

Potential Users run smaller operations on average, having less animals and land than Users and Nonusers. They are also slightly younger and less educated, but have the most experience. They have the least percentage of their income coming from ranching operations and have the smallest percentage of ranchers testing their bulls for Trich. Their attitudes toward risk are most similar to the Users group and they are the least suspicious of their neighbors' herd having been infected in the past 5 years. 89% get their information on the disease from their vet while 65% list as primary sources of information specialty magazines and other ranchers.

Nonusers have the largest herds but are second in land size, after Users. They have the lowest proportion of members using a PC and having internet available at home. Their dependence on ranch income is second to highest and on average they are the least familiar with the Trich treatment and the least risk averse. The primary sources of information are their herd veterinarian and other ranchers, just like the other two groups.

The factors included in the estimation procedure are discussed in detail in the Model Specification and Literature Background section of the paper.

3. Model Specification and Literature Background

Griliches'(1957) conclusion that economic attributes are at the root of adopting new technologies and innovations has received considerable attention and became areference point especially in agricultural economics research. The choice of exogenous variables to be included in the model is an important task when trying to explain the economics of adoption and diffusion of technology, and we turn to the existing literature for guidance. The current study investigates the adoption of the Trichomonaisis vaccine by ranchers in Nevada and is motivated by a puzzling finding: lagging vaccine adoption rates. Although the vaccine has been available on the market for the past twenty years, adoption rate are lower than expected, despite the fact that this venereal disease causes significant economic losses in the area. Trying to understand why ranchers hesitate to use available management practices, at the risk of incurring significant financial losses, is one of the objectives of the current study.

Since the survey used in this study was constructed in a similar fashion to the one used by Bhattacharyya et al (1997), we focus our attention on two sets of factors to be included in the adoption model: human endowment and production endowment. In addition, recognizing the importance of risk attitudes and risk perception in the adoption and diffusion of technology, we incorporate subjective measures of risk aversion and perception of the risk of exposure to the disease. Since the measures of risk aversion are self-reported, and thus intrinsic to the respondent, we consider them as part of the human endowment factors. A detailed discussion concerning the specific variables that are associated with human and production endowment follows below.

A widely recognized variable that characterizes human endowment is education. It facilitates information acquisition and understanding of new technology, and plays a key role in the speed of adoption and adjustment to change (Schultz, 1964) (Becker, 1993). Welch (1970) differentiates between worker ability and allocative ability, and finds that the allocative ability, or the ability to adapt to change, is positively related to education. The more educated farmers are, the faster they are able to adjust to changing farm and market conditions. In our adoption model, the education variable represents number of years of education. We also include a squared term of the variable in order to allow for a nonlinear relationship between years of education and the probability of adopting the vaccine. By doing so, we follow literature trends and are able to capture the diminishing returns to education.

Engle et al. (2006) investigate ambiguity aversion when it comes to the adoption of new technology in rural Peru. They find evidence that ambiguity aversion plays a decisive role in the adoption of modern crop varieties. Furthermore, Ruttan (1996) identifies the learning behavior of individuals as being an important factor in the adoption process. Thus, our econometric specification includes a variable that accounts for the degree of familiarity with the disease treatment, and can be seen as both a learning element and a measure of ambiguity. Respondents have to choose their level of familiarity on a scale from 1 to 5, where 1 is "not familiar" and 5 is "very familiar". Our a priori hypothesis is that familiarity with the treatment increases the likelihood of adopting the vaccine as it reduces the ambiguity surrounding new technology. This variable can also be viewed as a learning element. Since the vaccine has been marketed for 20 years, it is reasonable to believe that ranchers have some level of knowledge on this issue, especially since Trich has caused significant financial losses and has been a problem in the area. Familiarity and knowledge facilitate understanding and increases the probability of adopting the technology.

Another factor suggested in the literature as influencing the adoption and diffusion of technology is neighbors' attitudes toward it. Munshi (2004) looks at social learning in both homogeneous and heterogeneous populations and finds that technology adoption lags can at least partially be explained by the lack of prevalence of social information. Bauch and Bhattacharyya (2012) look at vaccine adoption and vaccine generated herd immunity, using evolutionary game theory, to capture the feedback between "disease prevalence and strategic individual vaccinating behavior". They address the free-rider problem when it comes to infectious diseases, as the unvaccinated individuals' risk of infection is reduced if a large enough proportion of the population is inoculated. Their results reinforce the importance of strategic social interaction and social learning in explaining changes in the risk perception associated with contracting the disease. In order to account for social attitudes regarding the Trichomonaisis disease, we construct variable *suspect neighbor* and *likelihood of exposure.* Variable *suspect neighbor* is a dichotomous variable equal to 1 if respondents answered yes to the following survey question, and zero otherwise: "Have you suspected or known that your neighbor's cattle were infected with Trich anytime during the past 5 years?" Taking into account the potential free-rider problem that may arise, we expect a positive association between this variable and vaccine adoption. If a rancher believes that her neighbor's cattle has been infected in the past five years, her perceived risk increases and thus she is more likely to adopt the vaccine and inoculate the herd. The variable that measures the likelihood of exposure is ordinal and responses range from “not at all” to “very likely”. Just as before, we expect a positive association between this variable and the probability of adopting the vaccine. The more likely it is that their cattle are exposed to neighbor’s cattle, the higher the likelihood of vaccinating.

By incorporating risk attitudes as part of the human endowment, the importance of risk aversion was recognized in the decision making process. As shown by O’Mara (1980), Binswanger et al.(1980) and others, risk preferences play an important role in the adoption of technological innovation. Kebede (1992) found that the degree of risk aversion played an important role in the adoption of new technology by Ethiopian farmers. Shapiro et al.(1992) (Shapiro, Brorsen, & Doster, 1992) included measures of risk aversion, using a Pratt-Arrow measure of risk attitudes. He found that adopters of double-cropping in the US, were more risk averse on average than non-adopters. Following the same trend, we include subjective measures of risk aversion as part of the human endowment factors. In our econometric specification, a risk aversion measure is constructed based on ranchers’ response when asked to describe their general willingness to take risk, measured on a scale from 1 to 5. Choosing 1 characterize the respondent as a “risk taker”, while choosing 5 is equivalent to saying that she “prefers to avoid risk”.

Recognizing that people might have different attitudes toward risk in the domain of gains than in that of losses (Kahneman & Tversky, Prospect Theory: An Analysis of Decision under Risk, 1979), the survey includes two lottery questions intended as objective measures of risk aversion. The first question evaluates risk attitudes in the domain of losses while the second one measures risk aversion in the domain of gains. Table 2 below summarizes the objective risk aversion measures by user category.

Table 9: Risk attitudes in the Domain of Losses and Gains

	% Users	% Potential User	% Nonuser	% Entire sample
Domain of losses				
Risk averse	82.76	79.03	70.27	78.34
Risk seeking	17.24	20.97	29.73	21.66
Domain of gains				
Risk averse— 1	32.26	46.88	43.59	40.61
-2	45.16	21.88	30.77	32.73
-3	11.29	14.06	5.13	10.91
Risk seeking 4	11.29	17.19	20.51	15.76

Survey results are consistent with Prospect Theory predictions in the domain of gains, as the largest proportion of respondents are risk averse. However, unlike theoretical predictions, the majority of the respondents are also risk averse in the domain of losses. The Users category has a slightly higher proportion of respondents being risk averse in the domain of losses and the lowest proportion of respondents being risk averse in the domain of gains.

In an effort to better contour risk attitudes, we also include variables that measure the tendency to worry about harmful, unplanned events. Four common concerns are enumerated and respondents are asked to choose between five alternatives that are organized in the form of a scale from 1,"never worry", to 5, "constantly worry". Participants are asked how much they worry about being hurt or killed in a car accident, being diagnosed with cancer, being bitten by a rattlesnake and getting struck by lightning.

In summary, seven variables are included in the model to capture human endowment: years of education, familiarity with Trich treatment, subjective risk attitudes, a scale that measures how suspicious respondents are of their neighbor's cattle having been infected in the past five years, a scale that measures the subjective likelihood of exposure to other rancher's cattle, how worried they are about harmful unplanned events and the level of concern about the Trichomonaisis disease.

We now move to discussing the production endowment factors and their role in the adoption of technology and innovation. Among these factors we include both farm and industry specific variables that characterize production in the area.

Feder (1980) and Feder and O'Mara (1982) find a positive association between technology adoption rates and land size. They claim that smaller agricultural producers are more risk averse, and thus slower at adopting new technologies, which they usually associate with a higher level of uncertainty and risk. Kobayashi and Tigran (2011) find a positive relation between the size of the operation, characterized by the number of animals, and the number and probability of biosecurity actions, in the context of a game between producers that participate in a livestock exhibition. Thus, empirical evidence suggests that factors that characterize the size of the ranching operation weigh heavily in the decision making process of ranchers when faced with the adoption of technology and biosecurity measures. Because of colinearity concerns, we limit the number of variables that indicate the size of the operation to two: income category and number of animals. *Animals* is a continuous variable that sums up the number of cows, heifers, bulls and steers. *Income* is a continuous variable obtained by using interval midpoint values. The survey asks ranchers to indicate the income interval that they pertain to out of the following 9 options: under $15,000; $15,000 - $24,999; $25,000- $34,999; $35,000- $49,999;$50,000- $74,999; $75,000- $99,999; $100,000- $149,000; $150,000-$199,999; $200,000 or more. Income is not only an indicator of the size of the operation, but also of the affordability of the vaccine. Specifically, as the number of animals increases, so does the herd inoculation cost. Thus, being in a higher income category facilitates vaccine adoption.

Region is a dichotomous variable that equals 1 if the ranch is located in the northern counties of the state, and zero otherwise. This is a region fixed effect that captures the difference between the operations located in different parts of the state. The main difference is that there is a lot more open range in the south and thus the risk of contracting Trich is higher and it is more difficult to implement other management practices such as fencing.

In an effort to better capture the characteristics of the operation, we also include an indicator variable that equals 1 if the respondent leases land for cattle grazing. This is an important characteristic because if they lease land for these purposes they might risk greater exposure to both feral bulls (this lands are usually not fenced) and other ranchers' cattle.

Because the legislation regarding Trich management practices is tailored differently by state, ranchers that transport cattle out of state will include this information in their decision making process. They might have a different set of costs and benefits than ranchers that have their operation confined within the state bounds. To capture these differences, we include variable *transport*, an indicator variable equal to 1 if ranchers transport cattle out of state and zero otherwise.

Ranch profitability over the next five years is a subjective measure and illustrates ranchers' optimism concerning their operation. We measure the degree of optimism by using a scale from 1 to 5. The higher the number chosen, the more optimistic ranchers are with regards to their future profits.

In summary, we include six variables in our model to characterize production capabilities and ranch characteristics: income, whether the ranch is located in the northern or southern part of the state, number of animals, if they lease land for cattle grazing, whether they transport cattle out of state and how optimistic the respondents are concerning their ranch's profitability over the next 5 years.

As mentioned earlier, three different adoption groups are examined. Those who answered "yes" to the following question belong in the Users category: "Do you ever vaccinate your herd against Trich?" Those who answered "no" are divided into 2 groups: Potential Users and Nonusers. If the respondents answered "yes" or "maybe" to vaccinating in the future, they are included in the Potential Users category[12]. Finally, the Nonusers category includes those participants that answered "no" when asked if they ever vaccinate against Trich and are not considering the vaccine for future use.

In our final model, out of 217 observations, 108 were used. This reduction in the number of observations is due to the fact that not all required information was available in the initial set.

4. Estimation Procedure

In order to analyze the adoption and diffusion of the Trich vaccine by Nevada ranchers, an adoption model is employed and the probabilities of respondents being in the Users, Potential Users and Nonusers categories are estimated. Since the dependent variable is categorical and has a natural

[12] The reason we have grouped these two responses is because only two people that do not vaccinate answered "yes" to doing it in the future. We believe that these two respondents are not that much different from the ones that answered "maybe" to the same question.

ordering (i.e. respondents in the Users category are closer to Potential Users than Nonusers), we start our analysis by using an ordered logit model to fit the survey data collected.

Following Greene (2010), respondents are assumed to have a certain level of unobserved utility, denoted $U^{*}{}_{ij}$, where i indicates the individual and j the alternative. U^{*}ranges from $-\infty$ to $+\infty$. Denoting individual's i choice with y_i, we can map the utility to choices as follows:

$y_i = 0$ if $-\infty < U^{*}{}_{ij} < \mu_1$

$y_i = 1$ if $\mu_1 < U^{*}{}_{ij} < \mu_2$

$y_i = 2$ if $\mu_2 < U^{*}{}_{ij} < +\infty$

In the inequalities above, μ_j represents threshold j. We have (J-1) number of thresholds, where J is the number of alternatives that the respondent has. In our specification, the respondents have three alternatives: they can be in the Users, Potential Users or Nonusers group. Thus, the number of estimated thresholds is equal to two. Based on the level of utility derived from adopting the vaccine, survey participants self-select in one of the aforementioned categories as follows: if the utility received from adopting the vaccine is smaller than a value μ_1, then the respondent is in the Nonuser category. Similarly, if the utility derived from using the vaccine is between some estimated values μ_1 and μ_2, the respondent is in the Potential User category, while any value larger than μ_2 places them in the User group. Thus, there is a "continuous range of preference" that are mapped based on the estimated thresholds to the discrete values that the dependent variable can take (Greene, 2010).

A general form of the random utility function for a survey participant is illustrated below:

$$U_{ij} = \beta_0 + \sum_{k=1}^{7} \beta_k x_{ik} + \sum_{m=1}^{7} \delta_m z_{im} + \epsilon_{ij}$$

The unobserved utility that individual i gets when choosing alternative j, U_{ij}, is modeled as a function of human and production endowment. Specifically, the vector of explanatory variables x, contains all the human endowment components, including risk attitudes. Similarly, vector z, includes the production endowment factors that might influence the adoption decision based on literature trends. The specification also includes a constant term, β_0, and the error term, ϵ_{ij}, that incorporates all the other factors, not explicitly modeled, that explain the level of utility individual i gets when choosing alternative j.

The probability that person i selects alternative j is thus given by the following expression:

$$p_{ij} = p(y_i = j) = p(\mu_{j-1} < U^*{}_i \leq \mu_j) = F\big(\mu_j - x_i{}'\beta - z_i{}'\delta\big) - F(\mu_{j-1} - x_i{}'\beta - z_i{}'\delta)$$

Where F is the logistic cumulative density function, $F(z) = e^{Z} / (1 + e^{Z})$.

Equivalently we can write $p_{ij} = p(y_i = j/x, z) = \frac{\exp(\alpha_j + x_i\beta + z_i\delta)}{1 + exp(\alpha_j + x_i\beta + z_i\delta)}, \alpha_j = -\mu_j$

In the ordered logistic regression, the coefficients of the explanatory variables do not vary by alternative (all the β and δ coefficients are the same regardless of the alternative chosen), but the intercepts do. The equality of the parameters between alternatives is known in the literature as the *parallel regression assumption* or the *proportional odds assumption*. Thus, this model assumes the same distance between adjacent categories of the dependent variable, when in reality this is rarely the case. If the estimated parameters are the same for all categories of the dependent variable, it implies the same effect of a change in the explanatory variable on all categories of the dependent variable. For example, this would imply that an additional year of education affects the probability of adopting the vaccine the same, regardless of whether the respondent is in the Users, Nonusers or Potential Users category. Another implicit assumption of this model is that the error variances are the same for all cases. If this assumption is violated, the standard errors are wrong and the parameter estimates are biased.

One of the solutions proposed in the literature is to use a multinomial regression instead, and thus obtain different parameter estimates by alternative, for all the explanatory variables. Taking into account the relatively small size of the data set used in the analysis, adopting this solution comes with the risk of losing not only statistical significance but also some information by ignoring the natural ordering of the alternatives. This approach is considered the least constrained as all the parameters differ by alternative.

An alternative solution is to use a model that does not assume proportionality and that allows the user to decide what parameters are fixed (constrained to be the same for each of the alternatives) and what parameters differ based on the categories of the dependent variable. This can be achieved by using either a generalized ordered logit model (Williams, 2006) or a heterogeneous choice model (Williams, Fitting heterogeneous choice models with oglm, 2010). However, this approach also has some shortfalls: internal inconsistency and not being able to ensure positive probabilities.

Although for theoretical reasons our ordered logistic regression seems appropriate, a series of tests are done in order to check if the parallel regression assumption holds. The results of five different tests are summarized in Table 3. The null hypothesis tested is that the estimated parameters are equal for the three categories of the dependent variable (Users, Potential Users, and Nonusers). If the null hypothesis is rejected then the specification is questionable and a less restrictive version might be more suitable.

Table 10: Tests of the Parallel Regression Assumption

	Chi2	**Df**	**p**
Wolfe Goulde	35.64	15	0.002
Brant	65.28	15	0.000
Score	36.57	15	0.001
Likelihood Ratio	46.08	15	0.000
Waldt	24.50	15	0.057

A significant p value means that we reject the null hypothesis that the vector of parameters is the same between the alternatives. Four out of the five tests indicate that the ordered logit model is too restrictive at 1% and 5% significance level, and all of the tests point to the same conclusion when using a 10% level of significance. Based on these results we consider a less restrictive approach keeping in mind the problems that might arise by doing so.

In order to relax the proportional odds assumption, we use a generalized ordered logit model that fits J-1 binary choice models separately and allows us to constrain the variables that meet the assumption, while allowing the ones that don't to be unconstrained. The ordered logit and the multinomial logit models are special cases of this more general model.

Following Williams (2006) the probability formulas for three special cases of the generalized ordered logit model are presented below:

Totally unconstrained generalized ordered logit model[13]:

$$P(Y_i > j) = \frac{\exp(\alpha_j + x_i\beta_j + z_i\delta_j)}{1 + exp(\alpha_j + x_i\beta_j + z_i\delta_j)}, j = 1,2, \ldots, J-1$$

Just as before, J is the number of categories of the ordered dependent variable. In this specification, both the intercept and vector of

13 This model is most similar to a multinomial logit model, where different parameter estimates are obtained for each category of the dependent variable.

parameters are allowed to vary by alternative. When j=2 the expression above is equivalent to a binary logistic regression. When $j > 2$ the generalized ordered model is equivalent to "a series of binary logistic regressions where categories of the dependent variables are combined" (Williams, 2006). Since we have 3 categories for the dependent variable, for j=0 we contrast the Nonusers with Potential Users and Users and for j=1, Nonusers and Potential Users will be contrasted with Users.

Totally constrained generalized ordered logit model[14]:

$$P(Y_i > j) = \frac{\exp(\alpha_j + x_i\beta + z_i\delta)}{1 + exp(\alpha_j + x_i\beta + z_i\delta)}, j = 1, 2, \dots, J - 1$$

This model is identical to the ordered logit model, presented at the beginning of this section. We allow for the intercept to be different but we restrict the parameter estimates to be identical for all alternatives.

Partially constrained generalized ordered logit model[15]:

$$P(Y_i > j) = \frac{\exp\left(\alpha_j + x_1\beta_1 + x_2\beta_{2j} + x_3\beta_{3j} + \cdots + z_1\delta_1 + z_2\delta_{2j} + z_3\delta_{3j} + \cdots\right)}{1 + \exp\left(\alpha_j + x_1\beta_1 + x_2\beta_{2j} + x_3\beta_{3j} + \cdots + z_1\delta_1 + z_2\delta_{2j} + z_3\delta_{3j} + \cdots\right)},\ j = 1, 2, \dots, J - 1$$

Also called partial proportional odds, this model allows for some of the coefficients to be alternative variant, while restricting others. In the general

14 This model is most similar to an ordered logit model, where the parameter estimates are the same for all categories of the dependent variable.

15 This model can be thought of as a combination of the other two. Specifically, some parameters will vary across alternatives while others will be fixed.

formulation above we restricted β_1 and δ_1 to be the same for all the alternatives while allowing β_2, β_3, δ_2, and δ_3 to be alternative specific.

In order to decide on the parameters that are allowed to vary by alternative and the ones that are constrained to be the same, separate Wald tests were run for each explanatory variable included in the adoption model. P values are summarized in Table 4 below. A significant test statistic indicates that the parameter violates the proportionality assumption and thus should vary by alternative.

Table 11: Parallel Lines Assumption Test of the Parameters

Parameter	p value
Income	0.2052
Region	0.2845
Animals	0.1063
Lease	0.3587
Familiar treatment	0.0010
Edu	0.8019
Edu2	0.1958
Transport	0.2005
Risk averse	0.0018
Suspect neighbor	0.0971
Worry	0.0675
Concern	0.5303
Profit outlook	0.1921
Likelihood of exposure	0.0016
Sum disease treated	0.3611

Results indicate that three independent variables violate the parallel line assumption: familiarity with the disease treatment (*Familiar treatment*), the subjective measure of risk aversion (*Risk averse*) and the likelihood of own cattle being exposed to other ranchers' herds (*Likelihood of exposure*).

Thus, the parameter estimates for these variables are allowed to vary by alternative while the rest of the parameters are fixed.

All three versions of the generalized ordered model presented above were estimated and compared and results are discussed in the following section of the paper.

5. Results

5.1. Qualitative Interpretation

Following the estimation procedure presented above and taking into account the violation of the parallel line assumption indicated by testing the model as a whole and each parameter individually, three models were estimated: a fully constrained generalized ordered logit model (equivalent to the ordered logit model, all parameters are fixed), a partially constrained generalized ordered logit model (some of the parameters vary by alternative, while others are constrained to be the same), and an unconstrained generalized ordered logit model (equivalent to a *multinomial logit* model, all of the estimated parameters vary by alternative).

Table 5 summarizes maximum likelihood estimates from the three different specifications. Qualitatively the parameters are the same in the majority of cases and affect respondents' choice in the same direction.

The fully constrained generalized ordered logit model produces one set of parameter estimates without differentiating by alternative. This is consistent with a uniform effect of a change in the exogenous variables on the dependent variable. We start by interpreting the parameter estimates from this specification and then explore the additional insights gained when relaxing the proportional odds assumption.

Table 12: Estimated Coefficients- Comparison Between the 3 Models

	Constrained	Partially	Unconstrained
Nonusers			
income	0.457*	0.369	0.772
region	-0.698	-0.966*	-2.028**
animals	-0.047**	-0.040*	-0.098**
lease land	-0.704*	-0.585	-1.573*
familiar w/trich treat	0.562***	0.238	0.348
edu	3.214*	4.889**	6.182*
edu2	-0.107*	-0.164**	-0.216*
transport	0.446	0.582	2.035**
risk averse	0.391*	1.016***	1.707***
suspect neighbor	0.871*	0.587	0.895
worry	-0.477*	-0.506*	0.154
concern	0.032	0.094	0.210
profit outlook	-0.458*	-0.718**	-1.566***
likelihood of exposure	0.670**	0.121	-0.209
number of diseases treated	0.109*	0.091	-0.005
constant	-30.648**	-39.792***	-48.934*
Potential Users			
income	0.457*	0.369	0.284
region	-0.698	-0.966*	-0.641
animals	-0.047**	-0.040*	-0.012
lease land	-0.704*	-0.585	-0.093
familiar w/trich treat	0.562***	1.531***	1.498***
edu	3.214*	4.889**	5.200**
edu2	-0.107*	-0.164**	-0.171*
transport	0.447	0.582	-0.290
risk averse	0.391*	-0.045	-0.362
suspect neighbor	0.871*	0.587	-0.048
worry	-0.477*	-0.506*	-0.897**
concern	0.032	0.094	-0.106
profit outlook	-0.458*	-0.718**	-0.429
likelihood of exposure	0.670**	1.352***	1.535***
number of disease treated	0.109*	0.091	0.130
constant	-33.064**	-47.880***	-50.295**
Users	base outcome	base	base outcome

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Results indicate that higher income and education, and being more familiar with the Trich treatment increase the likelihood of adopting the vaccine.

Transporting cattle out of state has a positive but statistically insignificant effect in the first specification. The positive effect might be explained by the fact that ranchers that transport cattle out of state have to comply with other states' regulation. Furthermore, these ranchers run bigger operations[16] and could suffer greater financial losses in case of a Trich infection. Thus, although it is costlier to vaccinate because of the larger herd size, the benefits of doing so outweigh the costs.

Subjective risk aversion increases the likelihood of vaccination, which is in agreement with most of the literature on adoption of new technology. The more risk averse ranchers perceive themselves, the more likely it is that they will vaccinate. Similarly, the more suspicious ranchers are of their neighbors having contracted Trich in the past five years, the more likely they are to be in a higher response category. The higher the perceived likelihood of exposure to other rancher's cattle, the higher the probability of adopting the vaccine. These results support the tragedy of the commons idea presented in the introduction. If a rancher suspects a neighbor's herd of being infected with the disease and believes there is a higher likelihood of exposure to the

[16] On average they have more animals than the ranchers who no not transport cattle out of state.

neighbor's cattle, she is more likely to adopt the vaccine in order to limit the incidence of the disease. Similarly, if she is not suspicious of the neighbor's cattle being infected and there is a low perceived likelihood of exposure to other ranchers' herds, she will be less likely to adopt the vaccine. Thus, a rancher's biosecurity actions will depend on beliefs concerning neighbors' behavior.

Variable *concern* was created using factor analysis in order to overcome issues of high correlation among survey variables with a similar pattern of response. The main assumption is that these variables are associated with an unobserved latent variable, or factor. Using factor analysis, the latent variable is estimated using the common variation in the observed variables. When constructing variable concern, we used highly correlated categorical variables that indicate the degree of concern regarding events such as: fluctuating cattle prices, drought, wildfire, herd contracting Trich, water rights, limited access to financing etc. Using factor analysis, all this information was used to estimate one continuous variable, *concern*, which was then included in the final econometric model. The higher the values of the estimate, the more concerned ranchers are about events such as the ones mentioned above, negatively impacting their activity in the next 5 years. The parameter estimate has a positive sign, which means that this variables moves in the same direction as the dependent variable. The more concerned respondents are the more likely they are to vaccinate. However, it is not statistically significant in any of the models presented.

A lot of the times, as part of their management practices, ranchers administer more than one vaccine at a time in order to minimize the costs associated with inoculating the herd. Thus, the positive sign of the coefficient associated with the number of diseases they treat or vaccinate against (*number of diseases*), makes intuitive sense: the more diseases they treat or vaccinate against, the more likely they are to also vaccinate against Trich.

Variable *region* is a dichotomous variable equal to 1 if the respondent's ranching operation is located in one of the northern counties. The negative sign of the estimate indicates that having the ranching operation in this area reduces the likelihood of adoption. This result aligns with prior expectations since there is less commingling of herds and less public lands used for grazing in the northern counties. Thus, interaction between herds is less likely to occur and the actual and perceived risk of cattle contracting the disease is smaller.

Since the risk of contracting the disease decreases only if all cows are inoculated, the more animals a rancher has, the more expensive it becomes to vaccinate. The costs associated with vaccination can become prohibitive, which can potentially explain the negative relationship between the size of the herd (given by variable *animals*) and the probability of adopting the Trich vaccine.

Lease is an indicator variable that equals 1 if the respondent leases land for cattle grazing and zero otherwise. It has a negative coefficient, so people that lease land are less likely to vaccinate.

The variable denoted as *worry* is negatively related to the dependent variable. This variable is the result of factor analysis and measures the respondent's tendency to worry about harmful unplanned events such as being hit by a car, being diagnosed with cancer, getting bitten by a rattle snake or getting struck by lightning. The negative sign of the estimated parameter tells us that the more people worry about unplanned events such as the ones previously mentioned, the less likely they are to be in the User category. A potential explanation might be that the vaccine is only efficient 65% of the times and individuals fear that their herd might still contract the disease even if they vaccinate. They might see this event as being a harmful unplanned occurrence in the same category as the ones mentioned above.

Variable *profit outlook* is a self-reported measure of the degree of optimism regarding ranch profitability over the next 5 years. The more optimistic respondents are, the less likely they are to vaccinate. Being optimistic about future ranch profitability might entail underestimating the probability of the herd contracting Trich and thus explains the negative relationship between this variable and the likelihood of adopting the vaccine.

The vector of parameters estimated using the partially constrained and the unconstrained generalized ordered model can be interpreted as coefficients from binary logit models where the categories of the dependent variable are collapsed into two groups. Since there are three categories of the outcome variable, the first panel compares the Nonusers category to Potential Users and Users grouped together. The second panel groups Nonusers and Potential Users and compares them to the Users category. This interpretation can be used for the explanatory variables that violate the proportional odds assumption and thus are allowed to vary by alternative. The parameter estimates that do not vary by alternative are interpreted the same way as the ones obtained from the fully constrained generalized ordered logit specification above (most similar to an odered logit).

Results in the first panel show that if the ranching operation is located in the northern counties and ranchers lease land for cattle grazing, respondents are less likely to be in the Users or Potential Users categories. The bigger the size of the ranch given by the number of animals, the more likely it is that respondents are in the Nonuser group. The more optimistic individuals are about their profits, the less likely they are to vaccinate or to even consider vaccinating in the future. All of these results are consistent with the results discussed above for the ordered logistic regression and they are significant at least at a 95% confidence level.

Education and risk aversion have a positive effect on the probability of adoption. The more educated and more risk averse individuals are, the more likely it is they are either Users or Potential Users. Transporting cattle out of state also has a positive effect on the likelihood of adoption and the results are significant at a 99% confidence level. The more suspicious ranchers are of their neighbors' cattle being exposed to Trich, the more likely they are to at least consider vaccinating in the future.

The second panel compares the Users category to Nonusers and Potential Users grouped together. Familiarity with the disease treatment and increased likelihood of their cattle being exposed to other ranchers' cattle are the main factors that increase the likelihood of being in the Users category. Being worried about unplanned events has a negative impact on the likelihood of being in the Users category and is the only statistically significant coefficient with a negative sign.

In the partially constrained generalized ordered model three coefficients vary by alternative: *familiarity with the treatment*, the degree of *risk aversion* and the *likelihood of their cattle being exposed* to other ranchers' cattle. The rest of the parameters are invariant and have the same signs and interpretation as the ones obtained from the fully constrained specification. The positive sign of the estimated parameter of the degree of familiarity with the treatment, indicates that the more familiar people are with the technology, the more likely it is that they will be in the Users or Potential Users group. The biggest effect of this variable is to move ranchers from Potential Users to Users. So if ranchers have not adopted the vaccine yet but are considering doing it in the future, more information on the treatment might persuade them to move towards adoption. Similar results can be observed for the likelihood of exposure of their cattle.

An interesting result is the change of the sign of the risk aversion coefficient. It has a positive sign in the first panel and a negative one in the second. Thus,risk averse people are more likely to be in the Potential User category than in the other two. The change of signs in the second panel suggests that risk averse ranchers tend to be indecisive about vaccinating[17] and have less extreme attitudes regarding the adoption of the vaccine. The biggest effect of risk aversion is to move ranchers from the Nonusers group into higher categories and it is one of the few factors in our model that has an effect on the respondents in this category.

5.2 Marginal effects and elasticities

The vector of parameters doesn't give information on the change in the estimated probability given by a change in one of the exogenous variables. Inorder to have a better understanding of the magnitude of the effect that a smallchange in the explanatory variables has on the probability of being in one ofthe adoption categories, the marginal effects and elasticities associated with a change in each of the variables, are calculated and summarized in Table 6.

Results indicate that a 1% increase in the years of education increases the probability of being in the User category between 0.73% and 0.97%,

---

[17] They are more likely to be in the Potential Users category, which includes people that are not vaccinating now but might do it in the future.

based on estimates from the constrained and partially constrained model respectively. Thus, the effect of education increases and is significantly higher in the partially constrained model. Familiarity with the treatment increases the probability of being in the User category by 0.30%. These results are statistically significant at a 99% confidence level. This effect is confirmed by the ologit model, although it is quantitatively smaller (0.13%). An increase of 1% in the likelihood of cattle exposure will also increase the probability of adopting the vaccine by 0.27% according to the restricted generalized ordered model and 0.15% based on the ordered logit model results. Alternatively, it decreases the probability of being in the potential adopters group by 0.27%. It does not have a statistically significant impact on the Nonusers category.

A 1% increase in the degree of optimism concerning ranch profitability, in- creases the probability of non-adoption by 0.10%, while it decreases the probability of using the vaccine by 0.14%. The likelihood of cattle exposure has a significant effect on adopters and potential adopters but it does not influence the Nonuser category. A 1% increase in the perceived exposure decreases the probability of being in the Potential Users group by 0.25% and it increases the probability of adoption by 0.27%. This results are highly significant and reinforced by the results of the fully constrained generalized order (ologit) model, although the effects are quantitatively smaller .

The subjective risk aversion only has a significant effect on the non-adopters. The partially constrained model indicates that an increase of 1% in risk aversion decreases the probability of being in the Nonuser group by 0.14% increases the probability of being in the potential users group by 0.15%. Although it is not statistically significant, the marginal effect of risk aversion on vaccine users is negative. This is an indication that more risk averse people will tend to be indecisive regarding the use of the vaccine. They might consider using it in the future but the associated risk of the treatment not being efficient might prevent them to move to a higher category.

Table 13: Marginal Effects of the Estimated Probability

dY/dX	Totally Constr.	Partially Constr.	Unconstrained
Nonusers			
income	-0.0569*	-0.0521	-0.0976
region	0.0788	0.1200**	0.2000***
animals	0.0058**	0.0057*	0.0124*
lease	0.0904	0.0842	0.2140*
fam w/trich trtm	-0.0700**	-0.0335	-0.0440
edu	-0.4000*	-0.6890**	-0.7820**
edu2	0.0133*	0.0231**	0.0273**
transport	-0.0553	-0.0815	-0.2590**
risk averse	-0.0487*	-0.1430***	-0.2160***
suspect neighbor	-0.1090*	-0.0828	-0.1140
worry	0.0593*	0.0713*	-0.0195
concern	-0.00398	-0.0132	-0.0265
profit outlook	0.0570*	0.1010**	0.1980***
likely exposure	-0.0834**	-0.0171	0.0264
sum diseases treated	-0.0135*	-0.0128	0.0007
Potential Users			
income	-0.0461	-0.0208	0.0368
region	0.0838	0.0850	-0.0574
animals	0.0047*	0.0023	-0.0098
lease	0.0656	0.0294	-0.1940
fam w/trich trtm	-0.0568*	-0.2690***	-0.2760***
edu	-0.3250	-0.2760	-0.3310
edu2	0.0108	0.0092	0.0092
transport	-0.0454	-0.0337	0.3200*
risk averse	-0.0395	0.1520***	0.2930***
suspect neighbor	-0.0855	-0.0329	0.1240
worry	0.0481	0.0285	0.2110**
concern	-0.00323	-0.00529	0.0493
profit outlook	0.0462	0.0405	-0.1060
likely cattle expo	-0.0676*	-0.2500***	-0.3550***
sum diseases treated	-0.0110	-0.0051	-0.0285
Users			
income	0.1030*	0.0729	0.0608
region	-0.1630	-0.2050*	-0.1430
animals	-0.0105**	-0.0079*	-0.0026
lease	-0.1560*	-0.1140	-0.0199
fam w/trich trtm	0.1270***	0.3020***	0.3210***
edu	0.7250*	0.9650**	1.1120**
edu2	-0.0242*	-0.0323**	-0.0365**
transport	0.1010	0.1150	-0.0617
risk averse	0.0882*	-0.0089	-0.0774
suspect neighbor	0.1940*	0.1160	-0.0102
worry	-0.1070*	-0.0999*	-0.1920**
concern	0.0072	0.0185	-0.0228
profit outlook	-0.1030*	-0.1420**	-0.0919
likely cattle expo	0.1510**	0.2670***	0.3280***
sum diseases trtd	0.0245*	0.0180	0.0278

***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

5.3 Goodness of fit

Predicted probabilities for each of the three groups, Users, Potential Users and Nonusers were calculated. The predicted probabilities were estimated using the following formula:

$$\hat{P}(j) = \sum_{i=1}^{N} \hat{P}_i(j)/N$$

Following Bhattacharyya et al., the predictive power of each of the three models is assessed by comparing the actual probabilities to the estimated ones for each category of the dependent variable. Table 7 below summarizes the results.

Table 14: Predicted and Actual Probabilities of Adoption

	Constrained	Unconstrained	Partially Constrained	
	Predicted $\hat{Pj}$	Predicted $\hat{Pj}$	Predicted $\hat{Pj}$	Actual
Users	0.3836 (0.2588)	0.4263 (0.3472)	0.3948 (0.3278)	0.3598
Potential	0.3940 (0.1376)	0.2854 (0.4920)	0.3587 (0.3301)	0.4021
Nonusers	0.2224 (0.2149)	0.2883 (0.3083)	0.2465 (0.2238)	0.2381

The completely unconstrained generalized ordered logit model (equivalent to mlogit) has the predicted probabilities that are farthest from the actual ones. The fully constrained model (equivalent to ologit, parameters do not vary by alternative) predicts conditional probabilities within 2.38% of the actual probabilities for the Users category, within 0.81%

for Potential Users and within 1.57% for Nonusers. Comparatively, the partially constrained generalized ordered model produces a better prediction than the completely constrained one only for the Nonusers group. For this category it gives a prediction within 0.84% of the actual probability.

A comparison of the goodness of fit of the models is done using the Likelihood Ratio test statistic. Following Williams (2006) a LR test is employed to confirm that the ordered logit is indeed too restrictive. A LR statistic of 46.08 and $prob > chi2 = 0.0001$ was obtained. The null hypothesis that the ordered logit model (ologit) is nested in the generalized ordered logit (gologit) was rejected and thus we can conclude that the ordered logit model (ologit) is too restrictive. This result is not too surprising since four out of five tests indicated that the model violated the parallel lines assumption. A second LR test is employed to determine if the partially proportional odds model is too restrictive. The LR statistics was 19.57 and $prob > chi2 =$ 0.0756. The hypothesis that the partially restricted model is nested in the unrestricted one cannot be rejected at the 10%confidence level. This result indicates that the partially restricted model obtained with the autofit option was not too restrictive and is more appropriate than the fully constrained model (ologit).

6. Concluding Remarks

The estimation of an ordinal logistic regression offers valuable information on the factors and characteristics of ranchers that influence their decision making process when it comes to adoption of technology. The most common critique of this model is that it often violates the implicit parallel line assumption and gives biased parameter estimates. In order to address this concern, two additional, less restrictive, models were estimated: an unconstrained generalized ordered logit model with alternative variant coefficients, and a partially constrained generalized ordered model where three of the parameters are unconstrained to vary by alternative.

When comparing the three models using a likelihood ratio test, both the completely restricted (ologit) and the partially constrained (gologit2) models were found to be too restrictive. On the other hand, the completely unrestrictive model (gologit) yields estimated probabilities that are the farthest from the ones observed in the data collected. Parallel results are presented from all three models and the vector of coefficients tells the same story regardless of the estimation procedure.

Three distinct categories of the dependent variable are identified and adoption trends of *T. foetus* vaccine are explained in terms of human capital, production endowment and risk attitudes.

Being in a higher income category has a positive effect on adoption, especially since the vaccine administration can become costly as the herd size increases and the grazing is on open land. Education, familiarity with the disease treatment, risk aversion and the number of diseases that ranchers treat or vaccinate for, increase the likelihood of adoption. As the likelihood of their cattle being exposed to other ranchers' cattle increases, respondents are more likely to be in the User group. Although it is not statistically significant in the econometric models, being concerned about negative events having an impact on ranch profitability has a positive effect on adoption. The subjective degree of risk aversion positively affects the likelihood of vaccinating. Allowing for the risk aversion coefficient to vary by alternative shows that being risk averse pushes respondents away from being in extreme categories (Users and Nonusers). Ranchers are more likely to at least consider adopting the vaccine in the future conditional on the degree of risk aversion. The biggest effect of risk attitudes is to make Nonusers move to the Potential Users category and consider future adoption of the vaccine.

The number of animals, leasing land for cattle grazing, and optimism regarding ranch profitability decrease the likelihood of adopting the *T. foetus* vaccine. Optimism with respects to their ranching activity combined with underestimation of the perceived risk of their herd contracting Trich, drives ranchers into the Nonusers category.

Previously conducted surveys have shown the importance of computer use, veterinary check-ups and cooperative extension programs in the adoption of new technology. It is interesting to find that factors that had a significant effect on adoption are now wide spread among ranchers but the adoption of the vaccine continues to be lower than desired. Although initially included in our estimation procedure, the use of internet and computers and regular veterinary checkups are not statistically significant in this study.

Although the vaccine has been available on the market for over 20 years, our results indicate that ranchers' familiarity with the treatment should be improved in order to increase the rates of adoption. It is one of the most significant factors that could influence ranchers to move from the Potential Users to the Users category. Survey results indicate the most effective channels of information to be specialty magazines and veterinarians. The majority of the respondents list these two as their primary sources of information, together with "other ranchers" in the area. Since the primary reason that respondents list for not vaccinating their herd is the belief of not having a Trich problem and not enough information on the vaccine, policy makers should focus their attention on diffusing information about the *T. foetus vaccine* and increasing awareness of the risk of contracting the disease. Information on the incidence of the disease in the area might help ranchers more accurately assess the risk of not adopting the necessary management practices.

The rates of adoption could also be improved by making some of the management practice, such as testing of bulls, mandatory. Our statistical analysis indicates that ranchers that transport cattle out of state are more likely to vaccinate, potentially because some of the neighboring states have made testing mandatory.

Lastly, anecdotal evidence paired with estimation results point toward the same conclusion: scientific work to improve the efficacy of the vaccine is needed to increase ranchers’ confidence in the technology and boost adoption rates.

References

Bauch, C. T., & Bhattacharyya, S. (2012). Evolutionary Game Theory and Social Learning Can Determine How Vaccine Scares Unfold. *PLOS Coputational Biology, 8*(4). doi:10.1371/journal.pcbi.1002452

Becker, G. (1993). *Human capital: a theretical and empirical analysis, with special references to education.* Chicago: The University of Chicago Press.

Bhattacharyya, A., Harris, T. R., Kvasnicka, W. G., & Veserat, G. M. (1997). Factors Influencing Rates of Adoption of Trichomoniasis Vaccine by Nevada Range Cattle Producers. *Journal of Agricultural and Resource Economics, 22* (1), 174-190.

Binswanger, H. P. (1981). Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India. *The Economic Journal, 91*(364), 867-890.

Engle-Warnick, J., Escobal, J., & Laszlo, S. (2006, September). Risk Preference, Ambiguity Aversion and Technology Choice: Experimental and Survey Evidence from Peru. *Working Paper*.

Feder, G. (1980, July). Firm Size, Risk Aversion and the Adoption of New Technology Under Uncertainty. *Oxford Economic Papers, New Series, 32*(2), 263-283.

Feder, G., & O'Mara, G. D. (1982). On Information and Innovation Diffusion:A Bayesian Approach. *American Journal of Agricultural Economics, 64*(1), 145-147.

Greene, W. H. (2010). *Econometric Analysis.*

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica, 25*(4), 501-522.

Kahneman, D., & Tversky, A. (1979, March). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.

Kebede, Y. (1992). Risk Behaviour and Agricultural Technologies: The Case of Producers in Central Highlands of Ethiopia. *Quarterly Journal of Internatioal Agriculture, 31*, 269-284.

Kobayashi, M., & Tigran, M. (2011). Strategic Incentives in Biosecurity actions: Theoretical and Empirical Analyses. *Journal of Agricultural and Resource Economics, 36*(2), 242-262.

Long, S. J. (1997). *Regression Models for Categorical and Limited Dependent Variables.* SAGE Publications.

Mahmoudi, M., Landis, M. T., Fadali, E. R., & Harris, T. R. (2013). *Nevada Trich Survey Results.* Technical Report, University of Nevada, Reno, University Center for Economic Development.

Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics, 73*, 185-213.

Ruttan, V. W. (1996). What Happened to Technology Adoption-Diffusion Research? *Sociologia Ruralis-Journal of European Society for Rural Sociology, 36*(1), 51-73. doi:10.1111/j.1467-9523.1996.tb00004.x

Schultz, T. W. (1964). *Transforming Traditional Agriculture*. New Haven: Yale University Press.

Shapiro, B., Brorsen, W. B., & Doster, H. D. (1992). Adoption of Double-Cropping Soybeans and Wheat. *Journal of Agricultural and Applied Economics, 24*(2), 33-40.

Thain, D., Bruce, B., & Torell, R. (2007a). Trich in Nevada. Reno, Nevada: University of Nevada Cooperative Extension.

Thain, D., Bruce, B., & Torell, R. (2007b). Trich in Nevada. What Other States Are Doing. Reno, Nevada: University of Nevada Cooperative Extension.

Thain, D., Bruce, B., & Torell, R. (2007c). Trich in Nevada. Producer's Concerns and the Financial Impact. Reno, Nevada: University of Nevada Cooperative Extension.

Welch, F. (1970). Education in Production. *Journal of Political Economy, 78*(1), 35-59. doi:10.1086/259599

Williams, R. (2006). Generalized ordered logit/ partial proportional odds models for ordinal dependent variables. *The Stata Journal, 6*(1), 58-82.

Williams, R. (2010). Fitting heterogeneous choice models with oglm. *The Stata Journal, 10*(4), 540-567.

Chapter 3

An Analysis of Firm's Relocation and Expansion Decision Using Survey Data

Abstract

Understanding what drives firms' relocation and expansion decisions is critical in formulating sustainable regional economic development policies. Using individual firm level survey data, the current study analyzes factors that influence past and future relocation and expansion decisions by combining the neo-classical, behavioral and institutional approach. Information on over 2,500 firms from different industrial classifications was collected, and both conventional, e.g. economic and physical infrastructure, and unconventional factors, e.g. quality of life and business climate indicators, were incorporated using a self-reporting framework.

Results point to internal, external and spatial factors as important predictors of the past and future relocation and expansion decisions and industrial sectors are compared.

1. Introduction

Understanding firms' relocation and expansion decisions and the main factors that shape the decision making process is critical in formulating regional economic development policies. Attracting and retaining firms that will not only increase the economic activity of an area but also serve as an incentive for other firms to relocate in the vicinity, represents one of the top priorities of local governments when it comes to economic development strategies.

The relocation and expansion of firms generates an increase in employment opportunities and government revenues by expanding overall economic activity through a multiplier effect that ripples through the local economy. Through this multiplier effect, the benefits that accrue to local municipalities expand beyond the direct increase in input demand and incorporate some indirect effects (Harris, Shonkwiler, Ebai, & Janson, 2000). For example, in his book, The New Geography of Jobs, Enrico Moretti (Moretti, 2012) talks about the benefits amassed by service workers in the brain hubs (e.g. Silicon Valley). The higher standard of living and salaries of these workers are an indirect effect of all the prosperous companies that clustered in the area.

Resource distribution and welfare are also associated with the firms' relocation decision and are important considerations in regional development planning (van Dijk & Pellenbarg, 2000) (Pellenbarg, van Wissen, & van Dijk, 2002). The dynamics of government's spatial policies and relocation decisions are important to understand and factors that influence firms must be examined. The out-migration and in-migration of firms must be considered simultaneously. Thus, attracting new firms and retaining existing ones are the two sides of the same coin. This can be considered a zero-sum game since when one firm relocates outside the region, that region's loss is another region's gain. Furthermore, the negative aspects of attracting new firms to an area must also be considered by local authorities when tailoring their regional economic development policies. Aspects such as pollution, employee policies and wages, the type of jobs they bring into the area (high paying or low paying) and whether the appropriate labor supply exists, involvement in the community and social responsibility are just a few factors that could be used in classifying firms into desirable or non-desirable.

The present study examines business expansion and relocation decisions for firms within the United States. Using a comprehensive business survey, we attempt to estimate firms' propensity to relocate or expand by looking at factors that were suggested in the literature to influence business location decisions. We differentiate between the decision to relocate and expand using a framework that encompasses both past and future preferences. The richness of our data set allows us to also investigate the footloose firms hypothesis, i.e. firms that have relocated in the past are more likely to relocate again in the future. Furthermore, the survey questionnaire also addresses the relative distance associated with relocation and expansion and the mental map idea introduced by the behavioral approach to firm relocation. As explained in Pallenbarg et al (2008), the larger the relocation distance the higher the uncertainty associated with the new location: “relocation to another geographical market may even be comparable to the inherent uncertainty of a start-up”. Following Van Dijk et al. (2000) this could reveal a firm’s desire to keep its employees and minimize their commute times. On the other hand, a firm’s decision to relocate farther away might be an indicator of changing economic conditions and incentives, market changes, firm’s life cycle etc. All of these hypotheses are discussed in further detail in the following section, which addresses the model specification based on previous literature.

The remainder of the paper is organized into five sections. We start with a brief discussion of previous work followed by a description of data and summary statistics. Next we present the methodology and results, followed by concluding remarks.

2. Literature review

Pellenbarg, van Wissen and van Dijk (2002) refer to the 1970’s as the “golden era of firm relocation studies”. Studies published in this decade especially in United Kingdom, Netherlands, Germany, France, Italy etc., focus on factors such as firm’s age, size and market to explain relocation decision (Brouwer, Mariotti, & van Ommeren, 2004) The following decade lacks studies that investigate the issue of firm relocation and this area of research regains attention only in the late 1990’s (Hu, Cox, & Harris, 2008).

Van Dijk et al. (2000) identify three different types of factors that firms take into consideration when they are making relocation or expansion decisions: pull factors, push factors and keep factors. Pull factors are those that attract a firm to relocate or expand in a certain location, push factors are those that influence the firm's decision to leave a certain location (such as expansion limitations) and keep factors are those that determine the firm to stay in the current location (such as large initial investments, large number of employees that require special training etc.) In contrast, location theory focuses mainly on pull factors which reflect the level of attractiveness of a site (Brouwer, Mariotti, & van Ommeren, 2004).

The theoretical framework of firm relocation encompasses three different approaches: neo-classical, behavioral and institutional approach (Machlup, 1967).

Van Dijk and Pellenbarg (2000) identify a correspondence between relocation trends and the economic cycle of regions. Their conclusion supports the theory according to which firms are profit maximizing agents that are constantly looking for ways to optimize their behavior. Changes in demand and supply determinants affect firm's location decision and are considered in order to take advantage of all information available on the market at any given point in time. This theory is along the lines of the *neo-classical approach* to firm relocation, which assumes that the firm is a rational agent that has perfect information on all relevant parameters. The neo-classical approach considers both *push* and *pull* factors as important determinants of firm location decision. *Push* factors are those that incentives the firm to leave its current location, while *pull* factors are those that attract the firm to a new location. Examples of push factors are expansion limitations that could prevent the firm from taking advantage of economies of scale, availability of workforce, increased real estate costs etc. Examples of pull factors are proximity to desirable natural resources, neighborhood and knowledge spillover effects, thick labor markets, proximity to customers or suppliers, advantages of agglomeration economies in a different location etc. Tax incentives are also considered pull factors although there is controversial evidence in the literature regarding their importance in a firm's decision to relocate, with advocates on both sides of the argument. For example, Deller et al (2016) look at manufacturing firms that move across state and find the marginal effect of the business climate, including monetary and non-monetary incentives, to be trivial. On the other side of the spectrum, Plaut and Pluta (1983) find tax incentives to be statistically significant in explaining regional growth. However, they find other determinants, such as factor markets and climate characteristics, more important.

The neo-classical approach presented above only considers firm external factors to be important determinants of firm's relocation decision. Another approach that has gained popularity, is the *behavioral location* theory. According to Brouwer et al. (2004), its novelty comes from recognizing the importance of firm internal factors[18] in the decision to expand or relocate and the information and rationality limitations it faces when making this decision. Thus the assumption of

---

[18] Examples of firm internal factors are the age of the firm, size, which phase of its life cycle it's in etc.

perfect rationality and information, embedded in the neo-classical approach, is no longer considered appropriate and the firm is bound to make sub-optimal choices. The decision to relocate or expands drifts from the decision of optimal location as it also considers firm's path dependence and takes into account both transaction and relocation costs. The costs implied by relocation can be quite significant and include not only the cost of gathering information and comparing different sites, but also the direct cost of dismantling and reassembly of equipment, moving of facilities and training new employees (McCann, Urban and Regional Economics, 2001) . "Mental maps" are also considered when a firm decides to relocate and according to Brouwer et al. (2004), if a firm decides it is optimal to do so, it will usually relocate near places that it is more familiar with than distant ones.

The *institutional approach* is the most encompassing out of the three and "it starts from the assumption that economic activity is socially and institutionally situated" (Brouwer et al. 2004). It recognizes the importance of "institutional factors" such as trust and co-operation, social and cultural values, particularly for larger firms that are more influential and thus have more negotiating power (Pellenbarg, van Wissen, & van Dijk, 2002).

The current study follows the institutional approach and uses a broad category of factors when analyzing the firm's relocation and expansion decision.

Solving the issue of the theoretical framework, we turn our attention to specific factors that most influence the firm's decision. Hu et al. (2008)

includes management structure and employment as internal factors, government policy, labor market issues and regional economic infrastructure as external factors, and market size and distance as location factors.

The relative importance of the three factor categories is controversial in the literature. For example, Schmenner (1980) and Mariotti (2005) conclude that internal factors have the highest contribution in a firm's relocation decision. Brouwer et al. (2004) finds that external growth factors are particularly important in explaining relocation behavior in large firms. The relevance of location factors is reinforced by findings of the studies conducted by De Bok and Sanders (2005) who conclude that distance to transportation facilities is the most important. Looking at firm relocation patterns in Europe, Pellenbarg et al. (2002) find that transportation plays a key role and that firms are more likely to relocate in the same municipality rather than farther away. These findings also support the "mental map" theory presented by McCann (2001).

Another popular debate in the literature is whether tax breaks and local incentives have a significant impact on firm's relocation decision. Greg Leroy (2005) in his book *The Great American Jobs Scam*, argues that subsidies offered by local governments in order to persuade firms to relocate in a certain area, are irrelevant to the firm's decision. Furthermore, he claims that the key factors that play a decisive role in the location decision are "affordable supply of key inputs and proximity to suppliers and customers". Labor constraints in particular play an increasingly important part as technical skills become essential and the Baby Boomers generation approaches mass retirement. Where the company is in its "life cycle" is another important aspect. Newer companies are more likely to cluster together in order to take advantage of agglomeration economies, while more established companies are more willing to be dispersed, as they shift their focus towards lower production costs. He makes the analogy of the prisoner's dilemma game because of the lack of cooperation between competing regions. Firms use this as an advantage that allows them to negotiate subsidies and increase the offer in their preferred location although they would have chosen to locate there regardless of the incentives package presented by local authorities.

Other studies also include some non-conventional factors in their analysis such as ethical factors, racial components, pollution control, quality of life, recreational demands of employees, environmental controls etc. Skiba (2006) looks at the proportion of immigrants in the workforce and conclude that it is an important factor to be considered in the location of production.

The current study incorporates a multitude of the aforementioned factors that are suggested in the literature as important relocation determinants and gives a comprehensive view of firm's decision to relocate or expand. Unlike the majority of previous studies that use macro level data to study the overall trend of firm relocation together with an economic impact analysis on the local economy (Hu, Cox, & Harris, 2008),the current study uses individual firm level data in an attempt to better understand both past and future relocation/expansion and distinguish results between the two-time frames. This study is most similar to the analysis done by Hu et al. (2008), as they both have the same starting point, survey data collected as part of the Area Sectoral Analysis Process (ASAP).

ASAP is a modeling tool used to help communities attain sustainable economic development by matching industries and communities based on desirability and compatibility indices. The process includes a business questionnaire that serves as the basis for constructing an industry profile that is matched to community goals and assets regarding three dimensions: economic, environmental and social. Business data for companies in different industries is being collected on an ongoing basis and thus the current study uses a more comprehensive data set compared to that of Hu et al (2008). The firm's surveyed in the more recent rounds, are moving away from fastest growing and highest paying industries and towards a more representative sample.

3. Data description and summary statistics

The data in this study has been collected across a thirteen year period, beginning in 2003 and ongoing as of 2016. There have been five rounds of survey collection. While the target population of U.S. firms representing specific four-digit NAICS sectors and a survey framework with a focus of understanding decision making around business relocation and expansion has remained predominantly constant, there have been changes in survey modes, sample sources, and refinement of variable collection in the survey design. Details regarding all questionnaire rounds that were used in the current study are provided below:

- Round 1: In 2003, a sample of 2,129 firms located throughout the U.S. was sourced from Dun and Bradstreet. The sample was stratified based on the top 100 fastest growing and highest paying of the 317 four-digit 2002 NAICS sectors (4NAICS) reported by the U.S. Bureau of Labor Statistics (all two-digit NAICS sectors were represented in the stratified sample). Round one surveys were collected by the Western Rural Development Center at Montana State University, Billings. Initially, invitations to participate by email were mailed to each firm, but due to low response rates the survey mode was changed to phone surveys part way through survey round one. Targeting five completed surveys from each 4NAICS, five firms were randomly selected and invited to participate. Firms declining to participate or not responding to relocation or expansion survey questions were replaced with another randomly selected firm from the same 4NAICS until the list was exhausted or the quota for that particular 4NAICS was met. A total of 213 of the 2,129 firms completed the survey in the first round resulting in a response rate of approximately 10% (Hu, Cox, & Harris, 2008).

- Round 2: In 2006 a second round of surveying occurred. Implementation of this round was identical in structure to the first round with a few exceptions which will be explicitly described here. The second unique stratified sample was sourced identically from Dunn and Bradstreet but contained 2,700 firms. The same 4NAICS quota procedure as round one was followed, but all firms in round 2 were contacted and participated solely by telephone. A total of 1,064 firms completed the survey in the second round resulting in a response rate of approximately 39% (Hu et al. 2008). A skip pattern coding error in the second round telephone survey instrument resulted in a large number of observations failing to be presented questions about current number of employees, percentage of full time employees, percentage of part time employees, desired acreage and square footage of facility if desiring to relocate or expand, as well as the percentage of square footage dedicated to retail, warehouse, office, and manufacturing within the desired facility square footage. A total of 746 of the 1,046 completed surveys in round two were impacted by the skip pattern error. In order to retain the effected 746 observations, values for the missing variables were imputed using regressions based on observed data from the Dun and Bradstreet database used to source the original sample.

- Round 3: A third round was implemented in 2008 using telephone interviews. Documentation of the sampling and implementation for this round is not available. The available dataset reflects 250 observations of which 14 are completed surveys. However, variables related to likelihood of relocating and or expanding appears to contain errors in skip patterns similar to round two, but applied less consistently. Due to the lack of historic documentation and inconsistently applied skip patterns, none of the data from round three is included in the final business database.
- Round 4: The fourth round includes changes in the goal/indicator definitions that resulted in modification to the structure of the business survey instrument. Two pilot surveys were implemented with the instrument modifications included and to determine an efficient solicitation method, and then the fourth round was conducted in three phases as follows:
    - Pilot one was conducted in fall 2009 by the Center for Research Design and Analysis (CRDA) at the University of Nevada, Reno. The sample targeted a 2008 list of 4NAICS reported as ‘green’ in a report by State of

California Employment Development Department (CEDD). A total of 200 firms were sourced from Dun and Bradstreet in a stratified manner similar to rounds one and two. These 200 firms were contacted by phone to request an email for survey invitation. Of the 35 firms providing an email, one firm completed the survey by email.

- Pilot two was conducted in winter 2010 by CRDA. This pilot targeted 100 firms and tested a mixed mode survey methodology of inviting firms by mailed letter to access an online link. A reminder letter along with a paper survey were mailed two weeks after the original letter. Of the 100 firms sampled, 25 were compiled in a stratified sample similar to rounds one and two and were driven by the CEDD list used for round four pilot one. The remaining 75 firms were compiled using the same sample stratification procedure but were sourced from Dun and Bradstreet based on a 2009 Pew Charitable Trust report titled “The Clean Energy Economy”. Six of the 100 firms completed the online survey; six of the remaining 94 firms completed the paper survey sent with the reminder letter.
- Phase one was conducted in summer 2010 and 1856 businesses were sampled from the Dun and Bradstreet data base. Both a conventional letter of invitation with a link to the online survey and a reminder letter with a paper version were sent to the entire sample. The response rate was approximately 3.18%.

  - Phase 2 was conducted during fall 2010 and 109 businesses were sampled from 2 separate lists provided by University Center for Economic Development (UCED) associates. Following the same methodology as before the return rate was 13.89%.
  - Phase 3 took place during winter 2011 and all surveys were conducted by phone.
- Round 5: The fifth round was conducted in fall 2013 and all surveys were done by telephone. This phase focused on Alternative Energy, Healthcare, High-Tech, Communication and Manufacturing NAICS codes. 80 four digit NAICS codes were selected and the required minimum number of observations for each of the NAICS codes was five. A sample of 150 firms was taken for each of the NAICS codes and the sample was pulled from the Reference USA data base. The business was either a headquarters, subsidiary or single location in order to avoid having multiple branches from one company. Three of the four digit NAICS codes did not have 150 businesses listed and thus the entire population was sampled.

All data obtained from the different survey rounds were compiled to ensure compatibility. The survey is divided in the following categories, encompassing the main areas of interest:

- Section 1: Relocation/Expansion History. This section focuses on past relocation and expansion. It looks at both reasons for relocating/expanding and spatial attributes such as distance from previous location and whether the move was within same city, county, state etc.

- Section 2: Physical Infrastructure. The importance of physical infrastructure is being investigated and factors that were considered or would be considered (for those firms that answered *no* to having relocated or expanded in the last five years) are ranked from *not all important* to *very important*. Factors include access to transportation networks, access to suppliers and customers, access to fiber optic lines, cell service, high volume water supply etc.
- Section 3: Economic Infrastructure. The importance of factors such as availability of managerial and skilled workforce, favorable local labor costs, incentives, short and long-term financing etc., were included in the analysis of economic determinants of relocation. Each factor's importance was measured on a scale from 1 to 4, where 1 is *not all important* and 4 is *very important*.
- Section 4: Quality of Life. Using the same scale as above, factors like low crime rate, availability of affordable housing and quality health care, quality of educational system, social and cultural opportunities, climate etc. were considered. These considerations are important, especially for headquarters if they want to attract and retain key executives that value cultural amenities and an overall high quality of life (Greg Leroy, *The Great American Jobs Scam*, 2005).

- Section 5: Information Sources. The use of different channels of information for making company relocation/expansion decision is investigated in this section. This information can be a starting point for policy makers that are trying to attract new firms in their areas as they will know which information sources are most effective.
- Section 6: Employee Benefits. An indicator of quality of jobs that a company offers, can be reflected in how a firm treats its non-managerial employees. Factors like paid vacations, health insurance, training programs and retirement plans with employer contribution are taken into account.
- Section 7: Company's Community Activities. The value of bringing a new company in an area stems not only from the number and quality of jobs offered but also from its impact on the community. Supporting cultural programs, health and wellness, youth athletic activities, local education and poverty alleviation shows that degree of involvement and care for the local community, and helps build a social responsibility profile of the firm
- Section 8: Business Relationships. This section investigates the benefits that firms get from agglomeration economies. Localization and urbanization economies are investigated in order to determine the economic climate that companies benefit most from. Attracting new firms in a region is just as important as insuring their success and prosperity. Proximity to suppliers, customers and employees are also

investigated to create a more accurate economic picture of the firm's needs when choosing a relocation/expansion location.

- Section 9: Future Relocation/Expansion. This section looks at firms' propensity to relocate or expand in the next five years. Proximity to previous location and factor inputs are investigated. The availability of land and facilities needed together with their use, expected number of employees at new location, level of skills needed and expected sales are also investigated. Information about expected number of employees at new location, level of professional skills needed and expected sales is also collected.

For each firm included in the survey, the data collected includes information on two types of decisions: a binary response for the past relocation and expansion, and an ordered response for future relocation and expansion decisions. The question regarding past relocation is formulates as follows: "Has your company relocated in the last five years?" Similarly, the question regarding past expansion is: "Has your company established additional locations in the last 5 years? "For both of these questions, respondents are given two alternatives: yes or no. The question referring to future propensity to relocate was formulated as follows: "How likely is it that your company will relocate in the next 5 years?" In the same fashion, the question referring to the future expansion was: "How likely is it that your company will establish an additional location in the next 5 years?" Respondents had to choose from 5 answer categories representing a Likert scale: *not at all likely, somewhat likely, likely, very likely* and *I don't know*. The responses recorded for this 4 variables are used as dependent variables of the econometric specifications presented in the next section.

Following Hu et al (2008), Table 1 gives the definition, mean or median and standard deviation of the variables of interest. Missing responses are replaced by the mean or median of the rest of the sample, which is an acceptable conservative approach when the proportion of missing data is relatively small (Hu et al.2008).

The variable with the highest median is accessibility to high speed internet, which makes it the most desirable feature in a firm's relocation decision. Access within 30 minutes to an interstate highway, as well as access within one day to customers and suppliers have median values equal to three[19]. Other important factors are the business climate (illustrated by favorable local business tax, favorable local labor costs and favorable worker's compensation tax rate), the availability of skilled workers, clean air and quality of education and healthcare. This is evidence that these factors are important considerations in the decision-making process. Accessibility to other means of transportation such as rail freight and ports are ranked least important with a median value of 1.

Previous studies have included dummy variables to indicate whether a firm belongs to the manufacturing sector or not, as the manufacturing industry is

[19] Responses to the importance of having access within 30 minutes to an interstate highway, were coded on a scale from 1 to 4, where 1 is "not important at all" and 4 is "very important". All of the infrastructure variables, including the importance of being within 1 day to both supplies and customers, are coded the same way.

often seen as an indicator of economic health (Reum Davis & Harris, 2006)(. The survey data used in the current analysis allows us to aggregate firms into several sectors and include fixed effect to see if companies from different sectors are more or less likely to relocate or expand. Although we have very detailed information about sectors (data was collected for 6 digits NAICS codes), we do not have enough observations in the data set to incorporate fixed effects at this level of disaggregation. Data was first aggregated using 2 digits NAICS sectors and then further into the following eight categories as explained below:

- Non-mobile=11+21+22+23
- manufacturing= 31 + 32 + 33
- trade= 42 + 44 + 45
- transport and warehouse = 48 + 49
- information = 51
- technology = 54
- health = 62
- other industries = 52 + 53 + 55 + 56 +71 +72 +81

Table 15 *Variable definition and descriptive statistics*

Variable	Definition	Mean/ Median	Std. Dev.
reloc*	has your firm relocated in the past 5 yrs	0.3786	0.4851
add loc*	has your firm expanded in the past 5 years	0.3681	0.4824
f reloc	how likely is it that your firm will reloc	1.0000	1.0737
f add loc	how likely is it that your firm will expand	1.0000	1.0759
Sales*	annual sales in millions	187.4791	2149.6368
Emptotal*	total employment in thousands	0.0621	0.3199
intstate	access to interstate	3.0000	1.1914
pkgfreit	access to package freight	2.0000	1.2225
railfreight	access to rail freight	1.0000	0.6798
pngerair	access to passenger air svcs	1.0000	1.0991
portharb	access to ports or harbors	1.0000	0.7330
supplies	access to supplies	3.0000	1.1061
customers	access to customers	3.0000	1.1801
thrphase	access to 3 phase electric power	3.0000	1.2734
natgas	access to natural gas	1.0000	1.0747
intntl trade	access to internatl trade port	1.0000	0.7252
fiberopt	access to fiber optic lines	2.0000	1.1985
hvolwhaters	access to high vol waters	1.0000	1.0262
hvolwastewater d	access to high vol wat disposal	1.0000	0.8882
solidwaste d	access to solid waste disp	1.0000	0.9953
highspeedinternet	access to high speed internet	4.0000	0.8888
localpubtrans	access to local public transp	1.0000	1.0635
futexpns atsite	possibility of fut expansion at site	2.0000	1.0881
manag	availability of manag workforce	2.0000	1.1148

skilled	availability of skilled workforce	3.0000	1.0920
technical	availability of technical workforce	2.0000	1.1496
unskilled	availability of unskilled workforce	2.0000	1.0352
llcost	favorable local labor cost	3.0000	1.0552
wtaxrate	favorable wage tax	3.0000	1.0527
loctax	favorable local business tax	3.0000	0.9759
training	availability of specialized training	1.0000	1.0170
financing	availability of financing	2.0000	1.1305
low crime	low crime rate	3.0000	0.8615
afford house	affordable housing	3.0000	1.0296
outdoor rec	outdoor recreation	2.0000	1.0739
clean airwat	clean air and water	3.0000	0.9349
shop op	shopping opportunities	2.0000	1.0329
cultural op	cultural opportunities	2.0000	1.0304
qual edu	quality of edu	3.0000	1.0942
qual healthc	quality of healthcare	3.0000	0.9328
public safe	public safety	3.0000	0.8839
non-mobile*	NAICS codes: 11+21+22+23	0.0735	0.2611
mfg*	NAICS codes: 31+32+33	0.3189	0.4662
trade*	NAICS codes: 42+44+45	0.2330	0.4228
transp wareh*	NAICS codes: 48+49	0.0643	0.2454
info*	NAICS codes: 51	0.0492	0.2162
tech*	NAICS codes: 54	0.1255	0.3314
health*	NAICS codes: 62	0.0476	0.2129
other ind*	NAICS codes: 52+53+55+56+71+72+81	0.0787	0.2694

* mean was calculated

Table 2 below presents summary statistics for each of the eight sectors defined above. We look at the mean or median value and standard deviation for each variable of interest. For the past relocation indicator, the *tech* sector (professional, scientific and technical services) has the highest mean with approximately 56% of the firms in this category having relocated in the past 5 years. It is followed by the *information* and *health care* (health care and social assistance) sectors. Not surprisingly, the lowest mean is observed in the *non-mobile* sector[20]. For the past expansion indicator, the *tech* sector also has the highest mean followed by *health care* and *information*. The lowest mean is observed in the *other* sector[21]. These findings indicate that firms from the *tech, information and health care* sectors were more likely to relocate or expand in the past five years.

Analyzing the future propensity to relocate, we see that seven out of the eight groups have a median value of 1. The only one that has a median of 2 is the *tech* sector. Thus, when looking at future expansion decisions, half of the firms in the groups say that it is not at all likely or somewhat likely to relocate in the next 5 years.

Mean sales are highest for the *other* sector, followed by *manufacturing* and *trade,* while the lowest values are observed for the non-mobile and health

---

[20] it includes agriculture, forestry, fishing and hunting, mining, quarrying, and oil and gas extraction, utilities, and construction.

[21] this group contains finance and insurance, real estate and rental and leasing, management of companies and enterprises, administrative and support and waste management and remediation services, arts, entertainment and recreation, and accommodation and food services

sectors. The *non-mobile* and *tech* sectors have the highest mean employment, of approximately 164 and 128 employees respectively. The lowest average number of employees is in the other sector and is approximately 23 employees.

Access within 30 minutes to an interstate highway is important to all of the eight groups, with a high value of 3 ("important") and a low of 2 ("somewhat important"). It is most important to the *manufacturing, trade, information* and *tech* sectors.

Access within one day, at a reasonable cost, to both suppliers and customers has a high median value across all sectors. Access to high speed internet has the highest median value just as it did when we looked at the entire sample. At the same time, access to a natural gas pipeline has the lowest median value across all groups. These results support the hypothesis that new technological infrastructures, such as high speed internet and 3-phase electric power, are becoming more important factors in firms' decision to relocate/expand while conventional infrastructures such as natural gas pipelines decrease in significance.

Availability of skilled workers and favorable local labor costs have a median value of 3 ("important") across all sectors and suggest the rising importance of these factors in a firm's decision to relocate/expand.

Table 16. *Variable characteristics by sector*

	Non-mobile		Manufacturing		Trade		Transport &warehouse	
Variable	Mean/Med	St. Dev.	Mean/Med	St. Dev.	Mean/Med	St. Dev.	Mean/Med	St. Dev.
reloc*	0.0874	0.2832	0.3245	0.4685	0.4671	0.4994	0.3938	0.4901
add loc*	0.1639	0.3712	0.2931	0.4555	0.4654	0.4992	0.3938	0.4901
f reloc	1.0000	0.7409	1.0000	1.0728	1.0000	1.0598	1.0000	1.1440
f add loc	1.0000	0.8980	1.0000	1.0598	1.0000	1.0549	2.0000	1.1366
Sales*	4.9902	24.978	268.8277	2461.562	205.3698	2597.363	19.6036	102.106
Emptotal*	0.0171	0.0569	0.0677	0.2372	0.0504	0.2193	0.0589	0.1807
intstate	2.0000	1.1451	3.0000	1.1419	3.0000	1.1555	2.0000	1.2654
pkgfreit	1.0000	1.0070	2.0000	1.1720	3.0000	1.2357	2.0000	1.2349
railfreight	1.0000	0.7031	1.0000	0.7461	1.0000	0.7269	1.0000	0.7215
pngerair	1.0000	0.8490	1.0000	1.0377	1.0000	1.0597	2.0000	1.2526
portharb	1.0000	0.7541	1.0000	0.7506	1.0000	0.8500	1.0000	0.8867
supplies	3.0000	1.0669	3.0000	1.0427	3.0000	1.1272	3.0000	1.1216
customers	3.0000	1.1433	3.0000	1.1648	3.0000	1.1162	3.0000	1.2191
thrphase	2.6275	1.2590	3.0000	1.1940	3.0000	1.2481	3.0000	1.2856
natgas	1.0000	0.9154	1.0000	1.1470	1.0000	1.1039	1.0000	1.1478
intntl trade	1.0000	0.6101	1.0000	0.7569	1.0000	0.8496	1.0000	0.8874
fiberopt	2.0000	1.2201	2.0000	1.1687	2.3964	1.1606	2.0000	1.2404

hvolwhaters	1.8326	1.2093	1.0000	1.0550	1.0000	0.9336	1.0000	0.9857
hvolwstwtr d	1.0000	0.9584	1.0000	0.8992	1.0000	0.8140	1.0000	0.9293
solidwaste d	1.0000	0.9914	1.0000	0.9711	1.0000	0.9209	1.0000	1.0501
internet	4.0000	0.9773	4.0000	0.8851	4.0000	0.9229	4.0000	1.0621
localpbtrans	1.0000	0.9785	1.0000	0.9760	1.0000	1.0198	1.0000	1.1432
futexpns	2.0000	1.1052	2.0000	1.0829	2.0000	1.0815	3.0000	1.0960
manag	2.0000	1.0816	2.0000	1.0883	3.0000	1.1071	2.0000	1.1250
skilled	3.0000	1.1135	3.0000	1.0605	3.0000	1.0951	3.0000	1.0621
technical	2.0000	1.1176	2.0000	1.1121	2.0000	1.1466	2.0000	1.1236
unskilled	2.0000	1.0477	2.0000	1.0731	2.0000	1.0180	2.0000	1.0034
llcost	3.0000	1.0619	3.0000	0.9623	3.0000	1.0148	3.0000	1.0840
wtaxrate	3.0000	1.0948	3.0000	1.0113	3.0000	1.0111	3.0000	0.9806
loctax	3.0000	0.9778	3.0000	0.9213	3.0000	0.9114	3.0000	1.0259
training	1.0000	1.0267	2.0000	1.0128	1.0000	1.0081	1.0000	0.9809
financing	2.1836	1.1673	2.0000	1.1295	2.3673	1.1163	3.0000	1.0986
low crime	3.0000	0.9607	3.0000	0.8583	3.0000	0.7937	3.0000	0.8625
afford house	3.0000	1.0460	3.0000	1.0054	3.0000	0.9896	3.0000	1.1126
outdoor rec	2.0000	1.0834	2.0000	1.0281	2.0000	1.0657	2.0000	1.1078
clean airwat	3.0000	0.8942	3.0000	0.9099	3.0000	0.9224	3.0000	0.9993
shop op	2.0000	0.9943	2.0000	1.0126	2.0000	1.0014	2.0000	1.0812

cultural op	2.0000	1.0600	2.0000	0.9916	2.0000	0.9888	2.0000	1.0501
qual edu	3.0000	1.1413	3.0000	1.0587	3.0000	1.0585	3.0000	1.1163
qual healthc	3.0000	0.9409	3.0000	0.9270	3.0000	0.8616	3.0000	0.9568
public safe	3.0000	0.9871	3.0000	0.8847	3.0000	0.7999	3.0000	0.8382

*mean was calculated

Table 2. Variable characteristics by sector

	Information		Prof. Scientific & Tech Svcs.		Health Care		Other	
Variable	Mean/Med	St. Dev.	Mean/Med	St. Dev.	Mean/Med	St. Dev.	Mean/Med	St. Dev.
reloc*	0.5041	0.5020	0.5623	0.4969	0.4706	0.5012	0.2030	0.4033
add loc*	0.4797	0.5016	0.5431	0.4989	0.5126	0.5020	0.1472	0.3552
f reloc	1.0000	1.1085	2.0000	1.1132	1.0000	1.1294	1.0000	1.0600
f add loc	1.0000	1.0867	2.0000	1.1155	2.0000	1.1323	1.0000	1.0594
Sales*	35.802	147.948	106.700	1269.764	6.729	24.343	737.344	4093.688
Emptotal*	0.0731	0.2345	0.1277	0.7148	0.0318	0.0926	0.0231	0.1494
intstate	3.0000	1.2740	3.0000	1.1825	2.0000	1.2467	2.0000	1.2011
pkgfreit	3.0000	1.2497	2.3268	1.2526	1.0000	1.2069	1.0000	1.0361
railfreight	1.0000	0.5680	1.0000	0.5258	1.0000	0.3520	1.0000	0.6162
pngerair	2.0000	1.1419	3.0000	1.1972	1.0000	1.0167	1.0000	0.9535

portharb	1.0000	0.6663	1.0000	0.5028	1.0000	0.4508	1.0000	0.5382
supplies	3.0000	1.0684	3.0000	1.1701	3.0000	1.1419	3.0000	1.1066
customers	3.0000	1.2485	3.0000	1.2089	3.0000	1.2081	3.0000	1.1807
thrphase	3.0000	1.2332	2.0000	1.2546	3.0000	1.2819	2.0000	1.2912
natgas	1.0000	0.9757	1.0000	0.9274	1.0000	0.9184	1.0000	0.9341
intntl trade	1.0000	0.6265	1.0000	0.5649	1.0000	0.1826	1.0000	0.5125
fiberopt	3.0000	1.1840	3.0000	1.1849	3.0000	1.2187	2.0000	1.2102
hvolwhaters	1.0000	0.8312	1.0000	0.8016	1.0000	1.1391	1.0000	1.1342
hvolwastewa ter d	1.0000	0.6724	1.0000	0.6641	1.0000	1.0361	1.0000	1.1039
solidwaste d	1.0000	0.9257	1.0000	0.8222	1.0000	1.2003	2.0000	1.1819
highspeedint ernet	4.0000	0.5968	4.0000	0.6696	4.0000	0.8901	4.0000	0.9085
localpubtrans	2.0000	1.1037	2.0000	1.0748	3.0000	1.2056	1.0000	1.1187
futexpns atsite	2.0000	1.1457	2.0000	1.0588	2.3464	1.0222	2.0000	1.0764
manag	2.0000	1.1970	3.0000	1.1031	3.0000	1.0640	2.0000	1.1458
skilled	3.0000	1.1688	3.0000	1.0752	3.0000	1.0126	3.0000	1.1314
technical	3.0000	1.1368	3.0000	1.0973	3.0000	1.1848	2.0000	1.1311
unskilled	1.0000	0.8026	1.0000	0.8466	1.0000	0.9974	2.0000	1.0555
llcost	3.0000	1.1396	3.0000	1.0819	3.0000	1.1582	3.0000	1.1513

wtaxrate	3.0000	1.1061	3.0000	1.0551	3.0000	0.9963	3.0000	1.1445
loctax	3.0000	1.0709	3.0000	0.9783	3.0000	1.0952	3.0000	1.0305
training	1.0000	0.9772	1.0000	0.9952	2.0000	1.1363	1.0000	1.0215
financing	2.0000	1.0924	2.0000	1.1044	2.0000	1.1044	2.0000	1.1816
low crime	3.0000	0.8404	3.0000	0.8703	4.0000	0.8104	4.0000	0.9525
afford house	3.0000	1.0562	3.0000	1.0148	3.0000	1.0316	3.0000	1.1164
outdoor rec	2.0000	1.1053	3.0000	1.0690	3.0000	1.1556	2.0000	1.1281
clean airwat	3.0000	0.9580	3.0000	0.9478	3.0000	0.9140	3.0000	0.9657
shop op	2.0000	1.1281	2.0000	0.9723	2.0000	1.0214	2.0000	1.1696
cultural op	3.0000	1.0423	3.0000	1.0020	3.0000	1.0616	2.0000	1.1088
qual edu	3.0000	1.1595	3.0000	1.0244	3.0000	1.1605	3.0000	1.2105
qual healthc	3.0000	1.0196	3.0000	0.8748	4.0000	0.7683	3.0000	1.0907
public safe	3.0000	0.9024	3.0000	0.9014	4.0000	0.7551	3.0000	0.9886

*mean was calculated

Most studies on firm relocation do not differentiate between the decision to relocate and expand. The data used for this analysis separates the two decisions and distinct models are used and compared throughout. In an attempt to better understand whether firms think of the concepts individually or collectively, we compare past decisions to relocate and expand in Table 3 below. In a similar fashion, Table 4 cross tabulates the future propensity to relocate with the future propensity to expand.

Table 17. ***Past relocation vs past expansion***

	Past Additional Location			
Past Reloc	Yes	No	No Response	Total
Yes	790	153	0	943
No	127	1,421	0	1,548
No Response	0	0	11	11
Total	917	1,574	11	2,502

Table 18. ***Future relocation vs future expansion***

	Future Additional Location					
Future Relocation	Not at all likely	somewhat likely	likely	very likely	No Response	Total
Not at all likely	1,247	118	35	41	16	1,457
Somewhat likely	52	357	10	16	4	439
Likely	15	8	228	11	0	262
Very likely	35	17	7	253	3	315
No Response	4	5	1	1	18	29
Total	1,353	505	281	322	41	2,502

We see that the highest proportion of firms that have answered *yes* to having relocated in the past five years have answered the same to having expanded in the same time frame. So, while it is clear that the majority of answers are consistent between the two decisions, it is not evident whether firms that have relocated have also expanded or whether respondents think about the decision to relocate and expand in a similar fashion. A similar pattern is observed when comparing future propensity to relocate and future propensity to expand. Most respondents give the same answers for both questions.

Table 5 aims to investigate the footloose firms hypothesis according to which firms that have moved in the past are more likely to move again in the future. Results indicate that out of 943 firms that have relocated in the past, 373 answer that it is not at all likely to relocate again in the next five years. Out of the 1,457 firms that said they were not at all likely to relocate in the future, about three quarters have not relocated in the past. On the other hand, among the 315 firms that said they were very likely to relocate in the future, 207 indicated *yes* to having relocated in the past. Thus our data shows some evidence in support of the footloose hypothesis, in line with previous findings in the literature (Hu et al. 2008).

In addition to the footloose firm theory we also investigate whether firms that established additional locations in the past are more likely to also expand in the future. Out of the 917 firms that have expanded in the past, approximately 64% say that they are at least *somewhat likely* to expand in the next 5 years, while 40% respond that they are *likely* or *very likely* to do so. Furthermore, among those that have not expanded in the past, the vast majority say that they will not do it in the future either. These findings support the idea that firms that have opened additional locations in the past are more likely to do so in the future as well.

Table 19. ***Past vs future relocation***

	Future Reloc.					
Past Relo	not at all likely	somewhat likely	likely	very likely	no resp	Total
Yes	373	187	168	207	8	943
No	1,081	249	93	105	20	1,548
No	3	3	1	3	1	11
Total	1,457	439	262	315	29	2,502

Table 20. ***Past vs future expansion***

	Fut add loc					
Past add loc	Not at all likely	somewhat likely	likely	very likely	No Response	Total
Yes	315	199	180	212	11	917
No	1,036	303	100	106	29	1,574
No	2	3	1	4	1	11
Total	1,353	505	281	322	41	2,502

Tables 7 through 10 below investigate the hypothesis that firms from different sectors act differently when it comes to both past and future relocation and expansion decisions.

All four dependent variables used in the estimation procedure are tabulated by sector. Table 7 shows the *info* and *tech* sectors have the highest proportion of firms that have relocated in the past. The *non-mobile* group followed by *other*, have the highest proportion of firms that have not relocated in the past.

Table 21**. *Past relocation by sector***

	non-mobile	mfg	trade	transp ware	info	tech	health	other	Obs.
Yes	8.74	32.45	46.71	39.38	50.41	56.23	47.06	20.30	943
No	91.26	67.55	53.29	60.63	49.59	43.77	52.94	79.70	1,548
Obs.	183	795	578	104	123	313	119	197	2,491

When it comes to establishing additional locations in the past, Table 8 indicates that *tech* and *health care industries* have the highest proportion of firms that gave an affirmative answer to this question. The highest percentage of firms that have not expanded in the past is seen in the *other* and *non-mobile,* followed by *manufacturing*.

Table 22***: Past expansion by sector***

Add loc	non-mobile	mfg	trade	transp_ wareh	info	tech	health	other	Obs.
Yes	16.39	29.31	46.54	39.38	47.97	54.31	51.26	14.72	917
No	83.61	70.69	53.46	60.63	52.03	45.69	48.74	85.28	1,574
Obs.	183	795	578	160	123	313	119	197	2,491

Table 9 looks at future propensity to relocate by sector. The *transportation and warehouse* industry, followed by the *tech* and *health care* sectors have the highest percentage of firms that are *very likely* to relocate in the future. This is an interesting result since the *tech* sector also has the highest proportion of firms that relocated in the past. This can be seen as further evidence in support of the footloose hypothesis mentioned above.

Table 23. ***Future relocation by sector***

Fut. Reloc	non-mobile	mfg	trade	transp wareh	info	tech	health	other	Obs.
Not at all likely	82.12	62.47	54.04	54.09	59.35	43.91	53.85	66.32	1,457
Somewhat likely	10.06	15.52	22.38	16.35	15.45	23.4	16.24	15.03	439
Likely	3.35	9.29	11.02	13.84	11.38	16.99	15.38	5.7	262
Very likely	4.47	12.72	12.56	15.72	13.82	15.71	14.53	12.95	315
Obs.	179	786	581	159	123	312	117	193	2,473

Table 10 shows that when asked about future expansion, firms from the *transportation and warehouse* sectors have the highest percentage that answer *very likely*. *Non-mobile, manufacturing* and *other* sectors have the highest percentage of firms that answer *Not all likely* to establishing new locations in the next five years.

Table 24. ***Future expansion by sector***

Fut. Expansion	non-mobile	mfg	trade	transp wareh	info	tech	health	other	Obs.
Not at all likely	67.96	60.64	52.6	44.23	55.28	43.27	43.97	60.42	1,353
Somewhat likely	18.78	17.95	23.18	21.15	21.14	23.72	21.55	19.79	505
Likely	6.08	8.97	11.94	17.95	9.76	16.99	18.1	6.77	281
Very likely	7.18	12.44	12.28	16.67	13.82	16.03	16.38	13.02	322
Obs.	181	780	578	156	123	312	116	192	2,461

4. Methodology

We employ two different models to analyze the data collected, a probit for the past relocation and expansion decisions, and an ordered probit for the future relocation and expansion decisions. The first model uses a binary outcome for the dependent variable that takes the value 1 if the firm surveyed has relocated or expanded in the last five years, and zero otherwise. The second model has an ordinal categorical variable as the dependent variable with four possible outcomes, ranging from *not at all likely* to *very likely*.

These two estimation procedures are commonly used in the literature that analyzes firm relocation or expansion decisions. Most studied employ either a logit or a probit model and both show consistent results. The main consideration when choosing between the two are theoretical. Following Hu et al (2008) we estimate a probit, and an ordered probit respectively, in the current analysis.

Following Greene (2010) (Greene, 2010), firms are assumed to get a certain level of unobserved utility associated with choosing each of the available alternatives, denote by U_{ij}^{*}, where i indicates the firm and j the alternative. In the probit model that has as a dependent variable whether the firm has relocated/expanded in the past five years, j is equal to 1 if the firm answered *yes* and zero otherwise. For the ordered probit model the number of alternatives, j, is equal to four. The unobserved level of utility U_{ij}^{*}, derived from choosing one of the alternatives, ranges from -∞ to +∞. Denoting firm i's choice with y_i, we can map utility to choices as follows:

$y_i = 0 \text{ if } -\infty < U^{*}{}_{ij} < \mu_1$

$y_i = 1 \text{ if } \mu_1 < U^{*}{}_{ij} < +\infty$

In the above inequalities, μ_j represents threshold j. We have (J-1) number of thresholds, where J is the number of alternatives that the respondent has. In the probit specifications, the respondent has only two alternatives (*yes* or *no*) and thus we only have one threshold. In the ordered probit models, the respondents have four alternatives (*not at all likely*, *somewhat likely*, *likely*, and *very likely*), which are equivalent to estimating three thresholds.

A general form of the random utility function for a participant firm is presented below:

$$U_{ij} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{ik} + \sum_{m=1}^{4} \delta_m z_{im} + \sum_{l=1}^{8} \theta_l w_{il} + \alpha_n + \epsilon_{ij}$$

In the above proposed empirical specification, the firm decision to relocate is a function of internal, external, and spatial factors. Specifically, the vector of explanatory variables x includes firm internal factors, the vector of parameters z includes firm external factors, while vector w includes spatial (location) factors. The specification also incorporates sector categorical variables to account for the sector specific differences among firms. The model also includes a constant term, β_0, and the error term, ϵ_{ij}, that incorporates all the other factors, not explicitly modeled, that explain the level of utility firm i gets when choosing alternative j.

5. Results

Taking into account the large number of survey questions, not surprisingly, we find that some of the variables are highly correlated to each other and thus we turn to factor analysis when deciding on the explanatory variables to be included in the model. We estimate an underlying latent factor, using the common variation in the observed variables with similar patterns of response, and then include the factor in the regression analysis as an explanatory variable. To decide whether factor analysis is appropriate, we use Cronbach's alpha (Cronbach, 1951) reliability coefficient. The reliability coefficient takes values between 0 and 1, where higher values are an indication that factor analysis is appropriate. Values between 0.6 and 0.8 or higher, indicate that factor analysis cab be used.

Using the method described above, we estimate 5 factors to be included in the econometric analysis as predictors of the relocation and expansion decisions: *rail, international port, high skill, local tax,* and *quality of life*.

Variable *rail* was constructed from two questions in the physical infrastructure section of the business survey: ”how important is it to have access to railhead or rail spur?” and “how important is access within 30 minutes to rail freight?”. Generally, companies that deemed one as important, also consider the other an important feature in the decision making process. It is reasonable to assume that the companies that consider rail freight important are those that ship a large volume of goods or that see it as an “emblem of innovation and a foundation for economic growth”, as described by Nick Lord in a recent article that was published in Politico Magazine (Lord, 2016).

Variable *international port* was constructed from two variables that summarize the answers to the following questions: “How important is access within 30 minutes to port/harbor facilities?” and “how important is access within 30 minutes to an international trade port?”. We assume that companies that rank harbor facilities and international trade ports as important have a common underlying feature, namely that they are international companies that need access to a port either for shipping products or for purchasing inputs.

The third variable estimated through factor analysis, high skill, was constructed using 4 survey questions that measure the importance of the availability of managerial, skilled and technical workforce, and the availability of specialized job training programs.

Local tax is a factor variable that is representative of the local business climate and was estimated using three variables: the importance of favorable local labor costs, favorable worker's compensation tax rate and favorable local business tax rates. Two other variables were initially included in the estimation of this factor variable, the importance of state and local government incentives, and of lenient environmental regulation. Since these two measures of local business climate were introduced in the later survey rounds, we had a lot of missing observations and decided to exclude them from the estimation procedure.

Finally, our fifth and last factor is a measure of the importance of *quality of life* indicators in a company's decision to relocate or expand. The following variables were used in constructing this factor: the importance of low crime rates, affordable housing, clean air and water, high quality natural ecosystem, outdoor recreation opportunities, social, cultural and shopping opportunities, quality of education and health care, access within 30 minutes to a college or university, and public safety. All of these measures were suggested in the literature as being important pull factors that could attract companies to an area by increasing the level of attractiveness of a site. These factors become especially important for companies that relocate or expand and want to retain their executive and management employees (LeRoy, 2005).

The sales volume, although important indicator of firm size, was highly correlated with number of employees so we decided to drop sales as we were also worried about endogeneity issues concerning this variable. Causality can occur from both directions, firms that have higher sales have a higher probability of relocating or, past relocation causes an increase in sales. Another reason to exclude this indicator from the analysis is that it has a lot of missing values, since sales information was not collected in the most recent rounds. However, we do include the number of employees at the actual site (in thousands) in order to account for the size of firm. Another firm internal factor that was considered is the age of the company. The firm's life cycle theory hypothesizes that younger firms are more likely to relocate in order to take advantage of agglomeration economies, as they are in the innovation stage where intellectual spillovers and large labor pools are more important than costs. Similarly, as the firm becomes older and more established, it focuses less on innovation and more on cost reduction and thus is less likely to relocate in order to take advantage of agglomeration economies. However, older firms could still relocate in order to take advantage of economies of scale or to take advantage of cost reductions in another location that is more profitable.

Spatial factors such as the proximity to an interstate highway and public transportation are also included in the estimation procedure, along with variables that indicate the quality of spatial surroundings, such as the quality of life indicator described above. Access within one day to suppliers and customers are also incorporated, as previous studies found a positive association between these two variables and the firm's decision to relocate (Hu, Cox, & Harris, 2008). Another location attribute included in the econometric model is the possibility of future expansion at site. This is an important element that has been hypothesized to weigh in the relocation and expansion decision since one of the reasons for relocation could be that the firm is growing and needs more space capacity. Thus, we expect a positive relationship between the decision to relocate or expand and the possibility of future expansion at site.

With the widespread use of internet services by firms in diverse industry, we also include the importance that respondents gave to infrastructure assets such as, fiber optic lines and high speed internet. We expect a positive association between the importance of these two infrastructure assets and the decision to relocate or expand.

Since some firms that are expanding or relocating might be in the beginning of their life cycle, or need financial support in order to pursue their expansion or relocation goals (might have large relocation or expansion fixed costs), the importance of available long and short term financing was also incorporated into the model. In the absence of financing, the relocation or expansion costs could be seen as keep factors that constraint the firm from relocating or expanding to a different site.

The availability of skilled and unskilled workers, are also important keep factors that should be considered in firm's relocation decision. Skilled labor might not be available everywhere and a company might have high hiring and training costs, making relocation costlier (van Dijk & Pellenbarg, 2000).

To summarize, we include three variables to account for firm internal factors: the number of employees, and the importance of the availability of skilled and unskilled workers. We include four explanatory variables that account for firm external factors: the importance of high speed internet, access to fiber optic lines, an indicator of the local business climate (*local tax*), and the availability of financing. Lastly, we include eight variables to account for spatial/location factors that influence the relocation and expansion decisions: access within 30 minutes to an interstate highway, access to suppliers and customers, the possibility of expansion at site, access to an international port and rail freight, availability of local public transportation and quality of life indicators.

For our main analysis we estimate four models, two probit and two ordered probit, using four different dependent variables and the same set of regressors. The probit models have as the dependent variable the firms' past decision of relocating or expanding respectively. On the other hand, the ordered probit models use the ordinal categorical variables regarding future relocation or future expansion. Tables 11 and 12 below summarize the results. Table 11 contains a comparison between the two probit models while Table 12 contrasts the two ordered probit models.

Table 25. Past relocation and expansion comparison–probit results

	Relocation ($N=2301$, $R^2=0.1459$)		Add location ($N=2301$ $R^2=0.1521$)	
Variable	Coefficient	p	Coefficient	p
Employees	1.256	0.000	1.421	0.000
Unskilled	-0.006	0.835	0.058	0.057
High_skill	0.187	0.000	0.222	0.000
Internet	0.089	0.018	0.090	0.016
Fiberopt	0.078	0.004	0.026	0.350
Local_tax	-0.009	0.837	-0.040	0.351
Financing	-0.030	0.298	-0.026	0.374
Qual_life	-0.041	0.257	-0.035	0.341
Intstate	0.116	0.000	0.106	0.000
Rail	-0.109	0.004	0.001	0.975
Int_port	0.108	0.006	0.102	0.010
Fut_expansion	0.179	0.000	0.154	0.000
Loc_pub_trans	-0.017	0.549	-0.066	0.026
Access_supp.	-0.039	0.187	-0.031	0.300
Access_cust.	0.033	0.230	0.044	0.108
Non-mobile	-1.002	0.000	-0.661	0.000
Mfg	-0.249	0.050	-0.373	0.004
Trade	0.196	0.127	0.165	0.202
Transp_wareh	-0.065	0.685	-0.092	0.563
Tech	0.328	0.017	0.322	0.019
Health	0.222	0.191	0.408	0.016
Other	-0.452	0.005	-0.809	0.000
Constant	-1.408	0.000	-1.297	0.000

The variables that represent factors internal to the firm are *employees, unskilled* and *high_skill*. The coefficient for the first one is positive and significant in both probit models and indicates that the number of employees positively influences the decision to relocate and expand. The availability of unskilled workers only has a statistically significant effect on firm's decision to expand. The coefficient estimate suggests a positive association between the availability of unskilled workers and firm's probability of establishing additional locations. . Although it is not statistically significant in the relocation model, it seems to indicate that availability of unskilled labor force increases the probability of expansion and decreases the probability of relocation. The last factor in this category, *high_skill*, reveals a positive and statistically significant relationship between the availability of high skilled workers and firm's probability to relocate and expand.

External factors that are included in the model are those that refer to government policy and contour the region's economic infrastructure. We find a positive association between access to technological assets, namely high speed internet and fiber optic lines, and the probability of relocating and expanding. These findings support the idea of increased importance given to technological infrastructure as opposed to the more conventional one, such as access to a natural gas pipe line. The other two external factors that we included in the estimation procedure, local business climate and long and short term financing options, have negative but statistically insignificant coefficient estimates. This finding supports the hypothesis predicated by Greg Leroy (2005) in his book *The Great American Jobs Scam*, that government incentives do not influence firms' location choice. Although the availability of long and short term financing does not have a statistically significant effect, its negative sign could be interpreted as evidence that firms that rank this factor high are less likely to relocate and expand. Although it might be counterintuitive, our finding is in line with other results in the literature (Hu et al. 2008). A possible explanation might be that firms who rank this factor high already have financing available at their existing location and it would be difficult for them to find a location with better financing opportunities. An alternative explanation could be that firms that rank financing as being important already have debt and do not want to take on the risk of relocating or adding new establishments.

The third category included in the estimation procedure refers to spatial or location factors. We find positive and significant coefficients for access to an interstate highway and international port in both models, indicating that these factors positively affect the probability of relocating or expanding. A positive relationship was also found between the decision to relocate or expand and the possibility of future onsite expansion. The coefficient of this factor is positive and highly significant in both probit estimations. Conversely, firms that rank as important more traditional factors such as access to rail freight and rail spur, were less likely to relocate. On the other hand, access to this more traditional type of infrastructure doesn’t have a statistically significant effect on the probability of expansion. Having access within 30 minutes to customers also increases the probability of relocation/expansion and the indicator for this variable has a positive sign in both models. On the other hand, firms that value having access to supplies are less likely to relocate or expand. These results are qualitatively consistent with previous findings in the literature, although they are not statistically significant in our model.

Looking at the decision to relocate and expand by industry group, we find that firms pertaining to the non-mobile, manufacturing, and other sectors are less likely to both relocate and expand compared to firms in the Information sector. Conversely, firms in the professional, scientific, and technical services (*tech*) are more likely to relocate and expand than those in the information sector. Lastly, firms in Health Care are more likely to expand but not more likely to relocate (the coefficient is positive but not statistically significant in the relocation model).

Overall, access to technology and infrastructure, together with the possibility of future on site expansion are the most important factors that firms take into consideration when deciding on relocation and expansion.

Table 26. Future relocation and expansion comparison–ordered probit results

	Future Relocation (N=2285, R^2=0.0688)		Future Add location (N=2274, R^2=0.0674)	
Variable	Coefficient	p	Coefficient	p
Employment	0.294	0.000	0.309	0.000
Unskilled	-0.002	0.937	0.057	0.031
High_skill	0.152	0.000	0.165	0.000
Internet	0.077	0.021	0.071	0.031
Fiberopt	0.045	0.066	0.050	0.035
Local_tax	0.037	0.326	0.053	0.157
Financing	-0.020	0.424	-0.007	0.786
Qual_life	-0.077	0.018	-0.095	0.003
Intstate	0.147	0.000	0.079	0.001
Rail	-0.052	0.108	-0.004	0.908
Int_port	0.052	0.132	0.065	0.056
Fut_expansion	0.195	0.000	0.199	0.000
Loc_pub_trans	0.024	0.347	0.023	0.358
Access_supp.	-0.041	0.117	-0.063	0.015
Access_cust.	0.046	0.060	0.067	0.005
Non-mobile	-0.391	0.013	-0.290	0.047
Mfg	0.079	0.499	-0.106	0.353
Trade	0.222	0.062	0.051	0.659
Transp_wareh	0.265	0.067	0.253	0.072
Tech	0.381	0.002	0.274	0.024
Health	0.273	0.080	0.327	0.030
Other	0.100	0.484	-0.004	0.976
cut 1	1.614	[1.223; 2.005]	1.371	[0.993; 1.749]
cut 2	2.173	[1.779; 2.567]	1.995	[1.614; 2.377]
cut 3	2.640	[2.243; 3.037]	2.492	[2.107; 2.877]

Table 12 compares results between the two ordered probit regressions looking at the future propensity to relocate or expand. The results indicate a generally consistent pattern with those displayed in Table 11 with only slight differences between variables that influence future propensity to relocate and expand, and those that were important considerations in the past decision concerning relocation and expansion.

Variables related to internal factors have the same signs and interpretation as in the above probit models. Specifically, the number of employees, as an indicator of firm size, has a positive and significant coefficient indicating that bigger firms are more likely to both relocate and expand in the future. Coefficient estimates for the importance of skilled and unskilled labor force are aligned with previous results. Namely, the availability of high skilled labor force increases a firm's future propensity to expand and relocate, while the availability of unskilled labor force positively affects only the future expansion decision.

Next, we analyze the importance of firm external factors. Just as in the previous estimation results, the availability of technological support, represented by high speed internet and fiber optic lines, increases the probability of adding new location or relocating in the next five years. An interesting finding is the change in the sign of the coefficient estimate of the local business climate. Although not statistically significant, it could be an indication of the importance of local tax incentives when it comes to future decisions to relocate or expand.

Looking at location factors, access within 30 minutes to an interstate highway and international port increase the probability of adding new location or relocating in the next five years. Just as before, the possibility of future on site expansion has a positive and highly significant effect on both types of future decisions. The proximity to suppliers and customers continues to be relevant and have the same causal direction as before. Furthermore, the coefficients associated with these factors become statistically significant when it comes to the future propensity to relocate and expand. Possibility of future expansion at site remains positive and significant.

The indicator variables for industry category point out that firms in the technical services and health care groups are more likely to relocate and expand in the future compared to those in the information category. The coefficient for trade becomes statistically significant and has a positive sign in the model looking at future relocation. Thus, firms in the trade sector are more likely to relocate in the future than those in the information category. The transportation and warehouse coefficient changes its sign in these specifications and becomes significant. Companies belonging to these group are less likely to open additional locations or relocate in the future. Finally, firms in the non-mobile category are less likely to expand or relocate in the future, just as expected.

6. Concluding Remarks

The importance of studying factors that are considered in firms' relocation and expansion is crucial to understanding when and how firms make these decisions. It has important implications for factor, labor and financial markets and is critical in understanding the impact of wealth distribution across different regions (Van Dijk and Pallenbarg, 2000).

In this study, data collected through survey methodology and a self-reporting framework were used. This allows firms' leaders to subjectively assess and rank a series of factors that may be relevant to their relocation and expansion decisions (Hu et al. 2008). One of the unique features of the study stems from the differentiation between past and future decisions to relocate and expand. Four econometric models were estimated in order to compare and contrast past and future relocation and expansion decisions: two probit models that use past relocation and expansion decisions as dependent variables, and two ordered probit models that use future propensity to relocate and expand as endogenous variables.

This estimation procedure allows for a direct comparison between the factors that influence past relocation and expansion and those that are important considerations in future decisions. The results are generally consistent between the models, indicating that past and future decisions are similar. Firm internal factors, such as the number of employees and the availability of skilled workers have a positive effect on both past and future decisions regarding relocation and expansion. On the other hand, estimation results suggest that the availability of unskilled workers only impacts the past and future decision to expand, and has no significant effect of the decision to relocate. Firm external factors, mainly high speed internet and fiber optic lines, positively affect firms' decisions to relocate and expand. An interesting finding was the change in the causal direction of the local business climate. A comparison between models points out that although companies do not see tax incentives as important considerations in the past decisions, they do show interest in local tax policies when considering future relocation and expansion. Having reasonable access to customers was also found to be an important consideration in the future decision making process.

Results show important implications for regional policy makers as well. Local governments can benefit from understanding firms' locational needs in order to formulate attraction and retention programs that spur the economic development of local areas. Understanding the industry sectors needs in terms of physical, economic and social infrastructure, can help tailor policies that incentivize desirable industries to at least considering relocating or expanding in the area.

References

Brouwer, A. E., Mariotti, I., & van Ommeren, J. N. (2004). The firm relocation decision: An emprirical investigation. *The Annals of Regional Science, 38*, 335-347. doi:10.1007/s00168-004-0198-5

Conroy, T., Deller, S. C., & Tsvetkova, A. (2016). Regional Business Climate and Interstate Manufacturing Relocation Decisions. *Regional Science and Urban Economics, 60,* 155-168.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

De Bok, M., & Sanders, F. (2005). Firm Relocation and Accesibility of Locations. Empirical Results from the Netherlands. *Transportation Research Record Journal of the Transportation Research Board*. doi:10.3141/1902-05

Greene, W. H. (2010). *Econometric Analysis.*

Harris, T. R., Shonkwiler, S. J., Ebai, G. E., & Janson, P. (2000). Application of BEA Economic Area in the Development of the Great Basin Fiscal Impact Model. *Journal of Regional Analysis and Policy, 30*(1), 77-94.

Hayter, R. (1997). *The Dynamics of Industrial Loaction: The Factory, the Firm and the Production sysytem.* Chichester: John Wiley and Sons.

Hu, W., Cox, L. J., & Harris, T. R. (2008). Understanding Firms' relocation and Expansion Decisions Using Self-REported Factor Importance Rating. *The Review of Regional studies, 38*(1), 67-88.

LeRoy, G. (2005). *The Great American Jobs Scam.* Berett-Koehler Publishers, Inc.

Lord, N. (2016, July). Freight rail effect: How one industry drives billions in economic activity. *Politico Magazine*. Retrieved from Politico.com: http://www.politico.com/sponsor-content/2016/07/freight-rail-effect

Machlup, F. (1967, March). Theories of the Firm: Marginlaist, Behavioral, Managerial. *The American Economic Review*, *57*(1), 1-33.

Mariotti, I. (2005). Firm relocation and regional policy. A focus on Italy, the Netherlands and the United Kingdom. Chapter 9 p213. Groningen: Groningen University.

McCann, P. (1995). Rethinking the economics of Location and Agglomeration. *Urban Studies, 32*(3), 563-577.

McCann, P. (2001). *Urban and Regional Economics.* Oxford: Oxford University Press.

Moretti, E. (2012). *The New Geography of Jobs.* Houghton Mifflin Harcourt.

Pellenbarg, P. H., van Wissen, L. J., & van Dijk, J. (2002). *Firm Relocation; State of the Art and Research prospects.* University of groningen. Systems, Organizations and Management Research Institute Research Report.

Plaut, T. R., & Pluta, J. E. (1983). Business Climate, Taxes and Expenditures, and State Industrial Growth in the United States. *Southern Economic Journal, 50*(1), 99-119.

Reum Davis, A., & Harris, T. R. (2006). Exploring Firm Location Beyond Simple Growth Models: A Double Hurdle Application. *The Journal of Regional Analysis & Policy*, *36*(1), 45-67.

Schmenner, R. (1982). *MAking Business Location Decisions.* Englewood Cliffs, NJ: Prentice Hall.

Skiba, A. (2006, March 31- april 1). Immigrationn, Firm Relocation and Wefare of Domestic Workers. *6th Annual Missouri Economics Conference Selected Papers*.

van Dijk, J., & Pellenbarg, P. H. (2000). Firm relocation decisions in the Netherlands: An ordered logit approach. *Papers in Regional Science*, *79*(2), 191-219.